

UNIVERSITE JEAN MONNET DE SAINT-ETIENNE
Groupe d'Analyse et de Théorie Economique Lyon Saint-Etienne

THESE
Pour obtenir le grade de
Docteur en Sciences Economiques
Présentée et soutenue publiquement par

David FRACHISSE

Le 14 mars 2011

**STRUCTURES ET DETERMINANTS DES COLLABORATIONS
AU SEIN DES PROGRAMMES CADRES DE RECHERCHE ET
DEVELOPPEMENT TECHNOLOGIQUE DE L'UNION
EUROPEENNE.
UNE PERSPECTIVE RESEAU.**

Directeurs de thèse:

Monsieur Pascal BILLAND, Maître de conférences, Habilité à diriger des recherches à l'Université Jean Monnet de Saint-Étienne.

Madame Nadine MASSARD, Professeur des Universités à l'Université Jean Monnet de Saint-Étienne.

Rapporteurs:

Monsieur Nicolas CARAYOL, Professeur des Universités à l'Université Montesquieu de Bordeaux.

Monsieur Koen FRENKEN, Professeur des Universités à l'Université d'Utrecht.

Suffragants:

Monsieur Pierre MOHNEN, Professeur des Universités à l'Université de Maastricht.

Monsieur Alain QUEVREUX, Chef du Service Europe de l'ANRT.

Remerciements

Mes remerciements sont destinés à toutes les personnes qui m’ont accompagné, aidé et encouragé au cours de ma thèse.

Je remercie mes directeurs de thèse, Nadine Massard et Pascal Billand, pour m’avoir permis de réaliser cette thèse sous leur direction. Merci pour votre disponibilité et pour vos nombreux conseils, remarques et critiques que vous m’avez distillés tout au long de ma thèse et qui m’ont permis d’avancer.

Je remercie l’équipe du CREUSET (aujourd’hui GATE LSE) pour son accueil et, plus particulièrement, Mr Bellet, ancien directeur du CREUSET, pour m’avoir offert l’opportunité d’intégrer ce laboratoire de recherche. Je remercie également la nouvelle direction, Mme Villeval et Mr Riou, d’avoir accepté mon affiliation au laboratoire pour cette année.

Je tiens à remercier plus particulièrement Corinne Autant-Bernard qui a su me donner des conseils avisés, qui m’a impliqué dans le projet IAREG et qui a accepté de relire ma thèse. Je remercie également Christophe Bravard qui m’a éclairé sur certains points qui pouvaient m’apparaître nébuleux et Gérard Chatagnon qui a contribué à ce que le traitement des bases de données se fasse dans la bonne humeur et sans crise de nerfs. Merci à Angélique Tholoniât pour les nombreuses discussions sur politiques d’innovation (pôles de compétitivité en particulier) qui m’ont servi pendant la thèse et qui me servent aujourd’hui dans l’exercice de mon métier. Merci à Nelly Exbrayat pour ses encouragements et pour la petite formation sur Stata (même si, au final, je ne m’en suis pas servi mais l’important n’est pas là). Merci à Charlène Laurent avec qui j’ai commencé à travailler sur PCRDT lorsque j’étais encore en DEA. Merci à Christine Largeron et à Bernard Kaddour pour le programme de désambiguïsation. De manière générale, merci à tous les membres du CREUSET, des plus anciens aux nouveaux en passant par les membres éphémères, qui par leur sympathie donnent envie de se rendre au laboratoire pour travailler.

Cette thèse n’aurait pas pu avoir lieu sans la région Rhône-Alpes et le cluster GOSPI. Ils m’ont permis grâce à leur soutien financier de pouvoir réaliser cette thèse dans de très bonnes conditions.

Pendant cette thèse, j’ai eu la charge pendant deux ans des travaux dirigés de statistiques. Ce fut une expérience très enrichissante et je remercie Gilles Panneaux de m’avoir fait confiance.

Je remercie mes collègues de bureau, Florent, Sylvie et Jérôme, pour leur patience pendant la dernière année.

Enfin, je remercie Nicolas Carayol et Koen Frenken qui me font l'honneur d'être mes rapporteurs ainsi que Pierre Mohnen et Alain Quévieux qui ont accepté de faire partie de mon jury.

Cette seconde partie des remerciements s'adressent à mes proches.

Je remercie mes parents qui m'ont soutenu tout au long de mon cursus scolaire. Ils m'ont donné l'envie d'apprendre et de me dépasser pour atteindre mes objectifs. Ils m'ont toujours encouragé et aidé. Ils m'ont laissé choisir mes orientations tout en m'épaulant et en faisant de leur mieux pour me mettre dans les meilleures conditions afin que je puisse m'épanouir dans mes études et dans ma vie de tous les jours. Je remercie également ma petite sœur Alexandra pour tous ses encouragements. Merci à ma mamie Antonia pour toute son affection et j'ai une pensée pour mon papi José.

La thèse fut aussi pour moi le bonheur de rencontrer ma future femme Angélique. Mon ange, je te remercie pour ton amour et toute l'affection dont tu as fais preuve jusqu'à maintenant, pour tes nombreux encouragements et ta motivation contagieuse. Je te remercie pour la relecture minutieuse de ma thèse et pour tous les conseils que tu as pu me prodiguer jusqu'à maintenant. Je remercie également ma future belle famille, Irène et André, pour leur soutien. Leurs encouragements répétés ont été une vraie source de motivation.

Il reste encore tellement de monde à remercier que je vais donc terminer par ceci. Merci à : Alexandre, Damien et Chloé, Jérôme, Loïc et Bernadette, Ludovic, Rémy, Stéphane et Anaïs, Thierry, l'équipe de foot du samedi, ma marraine Anne et son mari Gilles, Dominique et Christian, aux amis d'Angélique qui sont aujourd'hui les miens, à toutes les personnes de ma famille que je n'ai pas encore cité, à la famille d'Angélique (famille de "La Valla" et famille de "Saint-Just"), à Mr et Mme Mathieu,...

Encore une fois, merci à tous !

Table des matières

Introduction générale	12
I Structure et positionnement au sein des réseaux de collaborations en R&D appliqués aux PCRDT	33
Introduction	34
1 Définitions et apports statistiques de l'analyse des réseaux.	38
1.1 Notions de bases	39
1.2 Des données de collaboration aux réseaux : le choix d'un mode de représentation	40
1.2.1 "Two-mode networks" : une première étape dans la construction	41
1.2.2 "One-mode networks" : des réseaux centrés sur les collaborations entre agents	44
1.2.3 Hypergraphes : des représentations fidèles mais incompatibles avec l'analyse réseau	51
1.3 Propriétés générales des réseaux	52
1.3.1 Densité d'un réseau	52
1.3.2 Composantes et singletons	55
1.3.3 Distance et diamètre	56
1.3.4 Coefficient de cluster	59
1.4 Analyse des réseaux au niveau individuel	60
1.4.1 La pertinence des indicateurs de centralité les plus usités	61
1.4.2 Nouvelle approche de la centralité	78
1.5 La méthode cœur/périphérie	92
2 Les réseaux de collaborations des PCRDT	100
2.1 Des réseaux de collaborations individuelles aux propriétés favorables	102
2.1.1 Présentation des deux domaines technologiques	103

2.1.2	Les réseaux des PCRDT possèdent ils des propriétés favorables à l'innovation ?	110
2.1.3	Des agents aux rôles différents	131
2.2	Analyse temporelle des réseaux des PCRDT : application au sous-domaine des Nanotechnologies	144
2.2.1	Un réseau de plus en plus structuré	145
2.2.2	Un réseau plus robuste	147
2.2.3	Des agents de plus en plus impliqués	150
2.2.4	Des PME intégrées au réseau des nanotechnologies	153
2.3	Une agrégation des données au niveau régional pour compléter l'évaluation.	159
2.3.1	Les nouveaux États membres, des pays en retrait	160
2.3.2	Des réseaux de collaborations individuels aux réseaux de relations entre régions	162
2.3.3	Des régions fortement connectées	167
2.3.4	Le cœur géographique du réseau européen : la prépondérance des régions des pays dominants	169
2.3.5	Conformité des PCRDT à la géographie de l'innovation européenne	173
Conclusion		181
 II Déterminants des choix de collaborations au sein des PCRDT		
Introduction		187
 3 Étude des déterminants des choix de collaborations bilatéraux		193
3.1	Les données utilisées : prise en compte des effets réseaux grâce au 5 ^{ème} PCRDT	194
3.2	Le modèle économétrique	197
3.2.1	Spécificités des modèles économétriques appliqués aux réseaux de collaborations.	197
3.2.2	Quel modèle utiliser pour déterminer les facteurs influençant les choix de collaborations dans les PCRDT ?	199
3.3	Le rôle primordial des réseaux sociaux au sein des PCRDT et l'effet mitigé de la proximité géographique	202
3.3.1	Formulation des hypothèses testables : des réseaux sociaux à la proximité géographique	202

3.3.2	Premières estimations	215
3.3.3	La prise en compte des collaborations antérieures	223
3.4	Une hétérogénéité des comportements au sein des réseaux de collaborations.	230
3.4.1	Une propension à collaborer différente en fonction du type des agents du binôme	231
3.4.2	Des déterminants différents en fonction du type des agents du binôme	235
4	Expliquer l'intensité des liens	243
4.1	Présentation du modèle de Poisson	244
4.1.1	Distribution de l'intensité des collaborations	244
4.1.2	Présentation du modèle de Poisson	246
4.1.3	Des déterminants similaires pour expliquer l'existence et l'intensité des collaborations	248
4.2	Limites du modèle de Poisson et corrections	250
4.2.1	Test de surdispersion	250
4.2.2	Corrections de la surdispersion	252
4.3	Relâchement des hypothèses du modèle de Poisson	254
4.3.1	Spécification du modèle binomial négatif	254
4.3.2	Modèle de Poisson ou modèle binomial négatif ?	256
5	Dépasser les liens bilatéraux	260
5.1	Approche par les triades	261
5.1.1	Définition d'une triade	261
5.1.2	Construction des variables explicatives propres aux triades	263
5.2	Hypothèses du modèle et variables explicatives	265
5.2.1	Les déterminants traditionnels	265
5.2.2	Estimations des déterminants traditionnels	269
5.3	La prise en compte du contexte particulier des PCRD	273
5.3.1	"Triadic closure" : un rôle accru de la proximité sociale	273
5.3.2	Les agents des nouveaux États membres incontournables au sein des triades ?	276
	Conclusion	280
	Conclusion générale	287

Introduction générale

A partir des années 80, l'UE s'est fixée pour objectif essentiel d'améliorer la compétitivité des firmes européennes face aux firmes américaines et japonaises. Les Programmes Cadres de Recherche et Développement Technologique (PCRDT) créés en 1983 sont les principaux instruments dont disposent l'UE pour atteindre cet objectif. Dans le cadre de ces programmes, l'UE finance des projets de collaborations en R&D (Recherche et Développement) rattachés à diverses thématiques technologiques afin de favoriser l'innovation en Europe. Les agents se regroupent au sein de consortium et soumettent un projet à la Commission européenne qui décide ou non de le financer en partie. Les projets regroupent des agents de différentes nationalités, de différents statuts, puisque l'on retrouve parmi eux des firmes mais aussi des universités, des écoles, ... et de différents types, public ou privé. La volonté de l'UE est de faire de l'Europe et de ses firmes des leaders dans les domaines technologiques les plus avancés.

Cette politique des PCRDT marque une rupture avec les politiques industrielles des années 60 et 70, comme le souligne F. Rossi (2005, p10) : *“The research programmes launched since the 1980s were inherently different from the industrial policy programmes that were typical of the 1960s and 1970s, when the Community was trying to foster the creation of strategic industries, in line with individual member states’ efforts to promote “national champions”. In fact, the objective was not to create industries ex novo, but rather foster cooperation, innovation and commercialization.”* L'objectif des PCRDT est bien de favoriser la compétitivité des firmes européennes dans leur ensemble et non plus de quelques unes et ce par le biais de projets de collaboration.

Les PCRDT sont aussi le principal instrument dont disposent l'UE afin de mener sa politique d'innovation. A partir du sommet de Lisbonne (2000), l'UE souhaite la création d'un Espace Européen de la Recherche (EER). L'EER se veut un espace intégré dans lequel les agents européens collaborent de manière intensive dans l'objectif d'améliorer la recherche et l'innovation en Europe. L'UE ne souhaite pas atteindre cet objectif en se focalisant uniquement sur les principales puissances économiques européennes. Elle souhaite que l'activité d'innovation soit entreprise de façon harmonieuse au sein de son territoire de manière à ce que, parallèlement à l'accroissement la compétitivité des firmes européennes, il existe un effet de rattrapage qui permet aux pays européens moins développés d'apprendre et d'accélérer le processus d'innovation en leur sein. Notons que les différentes politiques menées par l'UE sont liées les unes aux autres. La politique d'innovation est notamment complémentaire de la politique régionale de l'UE (ou politique de cohésion) qui vise à un développement harmonieux et à obtenir une certaine cohésion

entre les États membres de l'UE. Le 6^{ème} PCRDT s'est ainsi déroulé dans un contexte d'adhésion de nouveaux États membres à l'UE.

Les PCRDT comportent un ensemble de dispositifs qui ont des missions diverses. Notre travail se focalise sur les collaborations qui se déroulent dans les priorités thématiques retenues dans le cadre du 5^{ème} PCRDT et 6^{ème} PCRDT. Ces priorités thématiques concentrent l'essentiel des financements attribués dans le cadre des PCRDT et représentent la plus grande partie des collaborations. Chaque thématique est décomposée en domaines technologiques eux-mêmes décomposés en sous-domaines technologiques. Lors de chaque programme cadre, les agents peuvent participer à différentes thématiques. Les thématiques sont transversales de telle manière que des agents issus de secteurs différents peuvent être amenés à collaborer sur une même technologie qu'ils exploiteront ensuite à des fins différentes. Les thématiques choisies sont confrontées aux mêmes problèmes que la recherche en générale, à savoir une recherche de plus en plus complexe, multidisciplinaire, avec des coûts élevés et qui demande de plus en plus d'avoir une masse critique suffisante pour la bonne conduite des projets (Commission Européenne, 2002). C'est dans cette optique que l'UE souhaite faire collaborer au sein de projets de R&D des agents d'horizons divers qui possèdent chacun des capacités propres.

Économie de la connaissance et PCRDT

La volonté de l'UE de construire un EER et de renforcer la compétitivité des firmes européennes, à partir des collaborations en R&D et de l'innovation, est inspirée des théories de l'économie de la connaissance.

Les propriétés de la connaissance

D. Foray (2000, 2004) souligne le rôle prépondérant joué par les collaborations dans la production de connaissances : *“Il est désormais largement admis que la connaissance produite pour l'innovation ne provient pas uniquement de processus de création qui seraient le fait d'individus isolés et même d'organisations fermées. Ces processus sont le plus souvent collectivement organisés par les industries et les réseaux de relations entre les firmes ainsi qu'entre les industries et la puissance publique.”* (D. Foray, 2004, p 43-44). Cassier et Foray (1999) distinguent deux principales formes de collaborations. La première renvoie *“à des mécanismes informels et spontanés d'échange et de partage de connaissances et de savoir-faire”* (Foray, 2004, p44-45). Cette forme prend naissance dans un cadre professionnel ou dans

un cadre territorial préexistant et porte en général sur la mise en commun de ressources déjà disponibles. Il n’y a pas alors d’engagement formel entre les agents, la collaboration est seulement basée sur la notion de réciprocité. La seconde forme de collaboration renvoie “aux processus formels de coopérations et d’apprentissage collectif” (Foray, 2004, p45). Cette deuxième forme est basée sur la notion d’accord qui régit la collaboration entre les agents. Les rôles des différents protagonistes sont alors clairement établis. Surtout, ces collaborations engendrent la création “d’espaces semi-privés de circulation et de mise en commun des savoirs qui sont éventuellement moins ouverts que les réseaux informels”. La seconde forme de collaboration mise en avant par Cassier et Foray (1999) correspond à la notion de collaborations telles qu’elles sont construites au sein des PCRDT. La création d’un espace semi-privé généré par la collaboration est un aspect essentiel de ce mécanisme de coopération entre agents qui trouve en grande partie sa justification dans la nature particulière du bien qu’est la connaissance.

La connaissance, qui est nécessaire à la production d’innovation, possède des propriétés particulières qui incitent à la création de collaborations et de réseaux. La première propriété est que la connaissance est souvent un bien non exclusif, c’est-à-dire que l’on ne peut pas empêcher les agents d’en profiter. C’est un bien fluide. Cette absence de contrôle total sur ce bien génère des externalités positives pour les tiers, c’est-à-dire que la connaissance produite par un agent bénéficie à d’autres agents sans que ces derniers n’aient à verser une compensation pécuniaire pour l’utilisation du bien. Il existe de nombreux canaux par lesquelles la connaissance peut se diffuser de façon incontrôlée. La seconde propriété est que la connaissance est un bien non rival. La consommation de la connaissance par un agent n’empêche pas la consommation de la connaissance par d’autres (Romer, 1993). La connaissance est également un bien cumulatif au sens où la connaissance génère elle-même de nouvelles connaissances. L’ensemble de ces propriétés font de la connaissance un bien particulier dont la production entraîne un rendement social fort.

Il est nécessaire de préciser que toutes les connaissances ne s’échangent et ne se diffusent pas aussi aisément (Foray, 2004). En effet, il existe un certain nombre de barrières à cette diffusion qui sont aussi bien contractuelles (contrat, cadre de la collaboration,...) que matérielles (tous les agents ne possèdent pas les matériaux et machines nécessaires à la mise en application des connaissances) ou cognitives et surtout il existe différentes formes de connaissances, toutes ne sont pas codifiées. Si l’information codifiée circule plus facilement, ce n’est pas le cas des connaissances tacites, des savoir-faire et des expériences pratiques. Ces connaissances se diffusent plus difficilement et demandent une certaine proximité pour pouvoir être acquises ou du moins pour pouvoir être exploitées. Ces connaissances, au rendement privé fort,

sont sources d'un rendement social moins élevé. Leurs propriétés permettent au contrôleur de la connaissance de mieux maîtriser sa diffusion. De la même manière le rendement privé de cette forme de connaissance augmente dans la mesure où il existe une phase transitoire entre la création de la connaissance tacite et le moment où celle-ci est codifiée, phase pendant laquelle le propriétaire de la connaissance peut tirer des rentes de monopole (Zucker et al., 1994). Les projets collaboratifs des PCRDT sont une des manières d'accéder, en partie tout du moins, à ces connaissances en collaborant avec les agents qui les détiennent. Le choix des partenaires au sein des PCRDT est donc un enjeu important.

Rendements privé et social de la connaissance

Le fort rendement social de la connaissance implique un dilemme pour l'agent producteur de connaissances puisque ce dernier sait que le rendement social sera supérieur au rendement privé qu'il pourra attendre de la production de connaissance. Dans un pays, l'État (l'UE à l'échelle européenne) a tout intérêt à ce que certaines connaissances soient produites. Pour autant, un agent individuel n'a pas forcément intérêt à investir dans la production de des connaissances dans la mesure où il ne peut capter l'ensemble des bénéfices découlant de cette production. On peut alors se trouver dans une situation où le niveau d'investissement privé est insuffisant du fait d'un défaut d'incitation. Par ses propriétés de non rivalité et non excluabilité, la connaissance est un bien public (D. Foray, 2004). Pour autant, reconnaître que la connaissance est un bien public ne signifie pas que l'État doit produire ce bien. Par contre, laisser faire le marché pourrait être source de production insuffisante du bien. On est alors dans une situation de dilemme social où l'État, dans notre cas l'UE, doit favoriser la production de connaissance et doit donc chercher à favoriser l'appropriation des bénéfices des connaissances par les agents pour les inciter à produire tout en ne freinant pas la diffusion et l'accumulation des connaissances.

Les PCRDT tel que conçus par l'UE peuvent apparaître comme une réponse à ce dilemme. Tout d'abord, l'UE finance en partie les projets de collaborations. Ce financement partiel permet, d'une part, de ne pas rendre la production de connaissance dépendante des seuls financements publics et, d'autre part, incite les agents à collaborer en diminuant les coûts auxquels ils font face. Par cette diminution des coûts, l'UE augmente le rendement privé des agents. De cette manière, elle rend plus attractive la production de connaissances pour les agents. Elle fournit de plus un cadre défini dans lequel vont se dérouler ces collaborations, cadre qui possède ses règles en matière de protection de la connaissance. Tout en favorisant les agents, l'UE s'assure

d'une production de connaissance d'un niveau suffisant lui permettant ainsi d'atteindre les objectifs qu'elle s'est fixée en termes d'innovation. De plus, les projets collaboratifs sont un moyen d'augmenter le rendement privé en diminuant les externalités de connaissance. Ainsi, pour D. Foray (2004, p66) : *“Dans le domaine de la production de connaissance, la création d'entités collectives (accord de R&D, centre technique, consortium de haute technologie) permet d'“internaliser les externalités” et donc d'atténuer le problème.”*. Autrement dit cette solution vise à réduire la taille des externalités en étendant le périmètre dans lequel les connaissances vont être volontairement partagées par tous les agents impliqués dans le projet (Foray, 2004).

Externalités de connaissance

Les externalités de connaissances sont un des éléments essentiels qui incitent les agents à collaborer. Les PCRDT sont ainsi un outil permettant à chaque agent de mieux contrôler la diffusion des connaissances qu'il produit mais, dans le même temps, les PCRDT sont des outils puissants de production d'externalités globales sources de rendement social.

Si un agent profite directement du savoir de ses collaborateurs au sein des projets, il peut aussi profiter indirectement des connaissances produites par ses collaborateurs dans d'autres projets qui impliquent des agents différents. De plus, les réseaux de collaborations du type PCRDT sont aussi un moyen de pouvoir apprendre plus rapidement pour un agent en collaborant avec des agents qui possèdent une importante expérience en matière de R&D et des connaissances variées. Les externalités de connaissances sont un élément primordial afin d'assurer l'efficacité du réseau.

Certains éléments favorisent la production de telles externalités. La mixité des agents source de diversité des connaissances est une condition favorable à l'apparition d'externalités. Une seconde condition favorable est la proximité technologique. Le fait que les agents possèdent des thèmes de recherche proches et donc des codes communs, même s'ils font une utilisation différente des avancées scientifiques et technologiques par la suite, favorise les externalités. De plus, de nombreuses études ont montré que la proximité géographique favorise les externalités. La proximité géographique et la proximité technologique sont en fait deux conditions qui permettent le développement de l'innovation (Audretsch et Feldamn, 2004). Toutefois, il est nécessaire de trouver l'équilibre en termes de proximité technologique car une trop forte proximité géographique couplée à une proximité technologique peuvent aussi conduire à une spécialisation spatiale qui nuit à l'innovation par l'absence de masse critique et/ou des effets de concurrence (Massard et Riou, 2002).

Les thématiques des PCRDT sont souvent transversales et couvrent dif-

férents secteurs d'activités. Ainsi une même technologie peut être employée à des fins différentes par des agents issus de secteurs différents. De cette manière, cela peut empêcher ou limiter les effets de concurrence. De plus, la diversité technologique peut aussi être source d'externalités grâce à des effets de complémentarité (Jaffe et al., 1993).

Vertus du travail collaboratif et PCRDT

Les théories mettent en avant la difficulté pour un agent de mener seul des travaux de R&D et les vertus du travail collaboratif dans un environnement économique qui mute de plus en plus rapidement. J.P. Archambault (2004) mettait en avant trois approches complémentaires reprises par A. Tholoniât (2008) dans le cadre de sa thèse sur les pôles de compétitivité : *“La première concerne l'innovation qui nécessite une réactivité et une qualité permanentes. La deuxième sous-tend la notion de réseau, le décloisonnement entre la sphère de la recherche et celle de la production étant souhaité, l'entreprise devient étendue. Enfin, la troisième insiste sur l'importance des externalités de connaissance dans le développement d'un territoire.”*

Dans chacune des trois approches décrites par Archambault (2004), on retrouve en partie l'essence des PCRDT. Les programmes cadres sont construits sur la base de projet de recherche dans le but de générer de l'innovation qui entraînera par la suite une hausse de la compétitivité des firmes européennes. Afin de générer l'innovation, les programmes cadres cherchent à attirer les agents les plus compétents en matière de R&D dans chacun des domaines financés et notamment les agents les plus avancés et qui disposent d'important moyens en R&D, que ce soit des moyens financiers et/ou humains. Les grandes entreprises ou les multinationales européennes leaders dans leur différents domaines sont alors visées mais pas seulement. En effet, les PCRDT favorisent la participation d'agents au profil différent car l'UE est consciente qu'un moyen de stimuler l'innovation est de faire appel à une mixité d'agents qui ont chacun leurs compétences propres notamment en mélangeant acteurs publics et privés de la recherche. Massard et al. (2003) soulignait l'importance de ces deux types d'agents qui mènent des recherches complémentaires et non substituables. La mixité des agents ne provient pas seulement de la différence public/privé mais aussi d'agents de nature différente qui conduisent une recherche différente mais complémentaire ou qui possèdent des capacités et des savoirs différents. Par exemple, il est intéressant d'associer des Petites et moyennes Entreprises (PME) à la démarche car ce sont des agents qui possèdent souvent une connaissance poussée dans un domaine particulier et qui ont des capacités de réaction fortes. De fait, il apparaît opportun d'associer l'ensemble de ces agents au sein de projets qui donnent naissance à

un réseau de collaborations en R&D. Dosi et Teece (1997) promouvaient le réseau comme la forme la plus pertinente d'organisation dans le but de créer de nouvelles connaissances et d'échanger des informations.

L'UE a construit les PCRDT de manière à obtenir des réseaux de collaborations en R&D générant de l'innovation. Les programmes cadres sont construits de manière à ce que les agents pris individuellement ont intérêt à participer aux PCRDT afin d'apprendre de nouvelles connaissances, de bénéficier des externalités propres à un réseau et d'obtenir des moyens supplémentaires notamment pour atteindre une masse critique. En favorisant la participation d'agents d'horizons divers les PCRDT s'assurent ainsi d'une base de connaissances et de moyens qui permettent par la suite de favoriser l'innovation et donc à l'UE de se rapprocher de son objectif.

Objectifs de la thèse

Il apparaît que les collaborations formées dans le cadre des PCRDT forment un réseau de collaboration. L'objectif de la thèse est de montrer de quelle manière l'analyse réseau permet de mieux appréhender les PCRDT. Nous utilisons les outils de cette analyse pour examiner les propriétés de la structure des collaborations formées dans le cadre des PCRDT et rendre compte des déterminants de collaboration dans le cadre de ces programmes.

Les PCRDT forment des réseaux de collaborations denses qu'il est difficile d'étudier seulement à l'aide d'une représentation graphique. Nous montrerons que l'analyse réseau possède un nombre conséquent d'outils qui permettent d'étudier de manière précise les réseaux formés par les PCRDT. L'analyse réseau permet notamment de montrer si les réseaux dans les PCRDT possèdent des propriétés favorables à l'innovation. Au sein de ces réseaux, il apparaît important de s'intéresser au rôle joué par les agents et chercher à identifier parmi ces agents ceux qui occupent des positions remarquables. L'analyse réseau est pluri-disciplinaire ce qui lui permet d'étudier différentes dimensions du réseau grâce à l'apport d'outils certes différents mais complémentaires. Afin de montrer que l'analyse réseau permet d'améliorer la compréhension des réseaux de collaboration en R&D dans les PCRDT, nous montrerons qu'il est possible sur la base des analyses menées d'évaluer en partie la réussite des PCRDT et de pouvoir réaliser certaines préconisations en termes de politique d'innovation.

S'il est nécessaire de s'intéresser à la structure du réseau, il est primordial de chercher à comprendre les comportements des agents et notamment leur incitation à collaborer puisque ce sont ces incitations qui vont déterminer par la suite la forme de ce réseau. Il est donc nécessaire de s'interroger sur

les déterminants des choix de collaborations. En effet, l'UE met en place un certain nombre d'incitations afin d'attirer les agents au sein des PCRDT. Il serait intéressant de connaître les déterminants des choix de collaborations afin de pouvoir améliorer l'efficacité de cette politique d'innovation en prenant en compte les motivations des agents pour accroître l'attractivité des PCRDT.

Analyse des réseaux sociaux et appréhension des PCRDT

Afin de se donner les moyens d'apprécier la capacité des réseaux à répondre aux objectifs fixés par l'UE aux PCRDT, il est nécessaire de produire des méthodes d'observation et d'analyse qui prennent en compte les effets sociaux au delà des effets observables au niveau des agents. L'analyse des réseaux sociaux est à cet égard intéressante, car elle offre des outils d'observation et d'analyse fondamentaux permettant d'apprécier les caractéristiques des réseaux. Un de ces outils est l'économétrie de réseaux qui permet de mettre en avant les incitations et les motivations des agents à former des collaborations.

L'analyse réseau est basée à l'origine sur la sociologie et sur la théorie des réseaux (l'étude des réseaux réels) et n'a cessé de se développer depuis les années 1950. La littérature sur l'analyse des réseaux sociaux identifie l'invention du sociogramme¹ par Moreno et Jennings (1938, 1945), qui marque le début de la sociométrie², comme précurseur de l'analyse des réseaux sociaux comme nous la connaissons aujourd'hui.

Définition de l'analyse des réseaux sociaux

L'analyse des réseaux sociaux s'intéresse aux liens qui existent entre un ensemble de nœuds qui sont souvent des acteurs ou des institutions. L'analyse des réseaux sociaux met au centre de l'analyse les relations entre les agents et non pas les caractéristiques de ces derniers. Les caractéristiques des agents sont importantes pour permettre par la suite d'expliquer plus finement les relations entre agents. Cette analyse propose des outils divers afin de rendre compte de ces relations et en cela est la méthode idéale pour étudier les PCRDT. Pour Wasserman et Faust (2007), l'analyse des réseaux sociaux : *“is far more than an intuitively appealing vocabulary, metaphor, or set of*

¹Un sociogramme est une représentation graphique des liens que possèdent un agent.

²La sociométrie est une méthode quantitative pour mesurer les relations sociales.

images for discussing social, behavioral, political, or economic relationships. Social network analysis provides a precise way to define important social concepts, a theoretical alternative to the assumption of independent social actors, and a framework for testing theories about structured social relationships." (Wasserman et Faust, 2007, p17). L'analyse réseau fournit ainsi des outils permettant de mesurer les relations entre agents. Enfin, elle permet d'étudier l'évolution des réseaux puisque ceux-ci sont souvent dynamiques.

L'analyse des réseaux sociaux utilise de nombreux concepts, outils et méthodes issus d'autres champs disciplinaires comme l'économie, les statistiques, l'économétrie, la biologie, . . . , qui peuvent être aussi bien qualitatifs que quantitatifs, et s'appliquent à des champs de plus en plus divers. "*Social network is inherently an interdisciplinary endeavor. The concepts of social network analysis developed out of a propitious meeting of social theory and application, with formal, mathematical, statistical and computing methodology.*" (Wasserman et Faust, 1997, p10).

En économie, l'analyse des réseaux sociaux a été utilisée afin d'étudier des réseaux de différente nature. Ainsi, Goyal, van der Leij et Moraga-Gonzalez (2006) s'intéressent aux réseaux de co-auteurs. Plus proche des réseaux de collaborations des PCRDT, on trouve une littérature conséquente sur les réseaux de collaboration en R&D que ce soit sur les brevets (Cloodt, Hagedoorn et Roijakkers, 2006), sur les alliances technologiques ou collaborations en R&D entre firmes (Burt, 1992 ; Gulati, 1998 ; Hagedoorn, 2002 ; Duysters et Hagedoorn, 2002 ; Verspagen et Duysters, 2004 ; Roijakkers et Hagedoorn, 2006 ; Deroïan, M'Chirgui et Milleli, 2007). Ces collaborations en R&D sont guidées par différentes motivations et sont aussi influencées par les politiques économiques (Hagedoorn, Link et Vonortas, 2000).

De nombreuses études portent sur les PCRDT et plus particulièrement sur la structure des réseaux des PCRDT et les agents les plus centraux des différents réseaux. Plusieurs études portent sur l'évaluation des réseaux (Malerba et al., 2006 ; Frachisse 2007) et cherchent à montrer si les réseaux des programmes cadres possèdent les propriétés adéquates (Breschi et Cusmano, 2002 ; Wagner, Thomson, Tesch et Perez, 2004 ; Roediger-Schluga et Barber, 2008) ou si les PCRDT ont réussi à attirer des agents leaders dans leur domaines. Le travail peut être réalisé en croisant différentes bases de données (Malerba, Vonortas, Breschi et Cassi, 2006). A l'exception du papier de Malerba et al. (2006) et de Frachisse (2007), les études menées concernent les 3^{ème}, 4^{ème} et 5^{ème} PCRDT.

Différents niveaux d'analyse

Si l'analyse des réseaux sociaux peut s'appliquer à des réseaux de nature différente, il existe aussi différents types d'études.

De nombreux articles réalisent une analyse du réseau au niveau macroéconomique (les propriétés générales du réseau). Ils cherchent à déterminer la structure du réseau et tentent d'identifier une forme particulière de réseau. La structure du réseau est importante car de nombreux articles montrent qu'elle affecte la performance des agents.

Granovetter (1973) fut le premier à pointer cette relation entre réseau et performance en relevant le rôle primordial joué par les liens "faibles" au sein des réseaux. Par liens faibles, il entendait des liens qui permettent de relier deux communautés ou deux sous-ensembles d'agents entre eux. A l'intérieur d'un sous-ensemble, les agents sont fortement connectés entre eux. Sans l'existence des liens faibles, ils ne seraient pas connectés avec les autres sous-ensemble d'agents. Dans un contexte d'innovation, les liens faibles permettent des échanges d'idées et de connaissances, ils sont donc particulièrement important pour favoriser la créativité. De plus, ces liens peuvent permettre aux agents d'appartenir à un réseau tel que tous les agents puissent communiquer directement ou indirectement. Sans ces liens le réseau aurait été éclaté. Notons que chez Granovetter, les liens faibles sont établis par des agents appartenant à des groupes denses qu'ils permettent de connecter. Burt (1992) note que dans certaines situations économiques, ce sont des firmes qui n'appartiennent à aucun groupe qui permettent de relier des groupes de firmes entre elles. Ces firmes comblent ce que Burt appelle des "structural holes". Burt (1992) a montré que les firmes, en comblant des "structural holes" entre des groupes denses, peuvent en tirer un avantage en termes de compétitivité. Relier différents groupes d'agents semble donc bénéfique à l'ensemble des agents et plus particulièrement à celui qui sert d'intermédiaire.

Si l'importance des liens faibles et des firmes comblant les "structural holes" est avérée, beaucoup d'auteurs mettent également en évidence l'intérêt des "small world network" ou réseau petit monde dans l'étude des réseaux de collaborations (Wagner, Thomson, Tesch et Perez, 2004 ; Goyal, van der Leij et Moraga-Gonzalez, 2006 ; Roediger-Schluga et Barber, 2008). Watts et Strogatz (1998) ont été les premiers à mettre en avant les propriétés de cette forme de réseau particulier. Le "small-world network" est caractérisé par une faible distance entre les agents, un coefficient de cluster élevé³ et le fait que la plupart des agents soient connectés directement ou indirectement⁴.

³Le coefficient de cluster mesure la densité des relations entre les partenaires d'un agent.

⁴L'ensemble de ces concepts seront définis dans la partie 1 de la thèse.

Cette forme de réseau est particulièrement recherchée dans les réseaux de collaborations en R&D, car elle possède des propriétés favorables à la performance des agents et à l'innovation. Dans une étude réalisée à partir de simulations, Cowan et Jonard (2004) ont montré que les "small world networks" obtenaient de meilleures performances que des réseaux aléatoires ou des réseaux plus réguliers en terme de distribution des liens⁵ en matière de diffusion des connaissances. Il existe de nombreux travaux qui ont montré les liens entre structure du réseau et performance. Parmi ces travaux, on peut citer les articles de Burt (1994) sur les liens entre réseaux sociaux et performance individuelle, Hagerstrand (1969) et Coleman (1966) sur la diffusion des technologies, ... Certains articles nuancent toutefois l'idée d'une relation entre la structure du réseau et la performance en mettant en avant le fait que si la structure du réseau influence la performance c'est qu'il existe également des éléments extérieurs qui interviennent (Sandström et Carlson, 2008 ; Ozman, 2009). Le contexte, les pays et leurs normes mais aussi les caractéristiques des agents qui y participent agissent aussi sur le résultat final. De plus, ces articles montrent aussi qu'il est parfois difficile d'établir des lois générales dans la mesure où la plupart des travaux empiriques portent sur des sujets et des données différents. On sait que l'UE cherche à attirer dans les PCRD des acteurs dominants de la recherche comme les grands centres de recherche et les grands groupes. De plus, elle opère une sélection sur les projets financés puisque tous les projets soumis ne donnent pas lieu à un financement. De cette manière, elle cherche à s'assurer que les conditions extérieures notamment la qualité des agents soient présentes. Combiné à sa volonté de construire un EER intégré dans lequel les agents interagissent, on note que l'UE cherche à créer le cadre idéal pour l'innovation.

De manière complémentaire à l'examen de la structure du réseau, il est intéressant d'étudier l'évolution du réseau et de sa structure. De nombreuses études s'intéressent à ce point. Ainsi, Goyal, van der Leij et Moraga-Gonzalez (2006) mettent en avant le fait que le réseau de co-auteurs en économie sur la période 1970-2000 tend à se densifier et, dans le même temps, montrent qu'il possède au final les propriétés d'un small-world network.

Pour ce qui concerne les collaborations entre les firmes, l'analyse des réseaux permet ainsi de montrer leurs évolutions dans le temps. On peut se référer au travail de Hagedoorn (2002) de manière générale ou de Hagedoorn et Roijakkers (2003) dans le domaine de la biotechnologie et de la pharmacie.

⁵Une des propriétés des "small world networks" est la présence de "hubs", c'est-à-dire d'agents qui possèdent de nombreux liens et qui rendent la distribution des liens entre les agents au sein du réseau très inégalitaire.

L'analyse des réseaux sociaux permet ainsi d'obtenir des indications chiffrées de l'évolution des réseaux que l'on peut ensuite mettre en relation avec le contexte dans lequel se déroule ces collaborations, comme l'on fait Deroïan, M'Chirgui et Milleli (2007) sur trois industries high-tech.

Un deuxième niveau d'analyse, après le niveau macroéconomique, est le niveau microéconomique qui s'intéresse au rôle des agents au sein des réseaux et à leur évolution au sein de ce réseau. L'analyse des réseaux sociaux permet d'identifier au sein des réseaux des agents qui occupent une position particulière. Pour cela, l'analyse des réseaux fournit un certain nombre d'indicateurs qui permettent d'obtenir une mesure de la place occupée par l'agent dans le réseau en fonction d'un certain nombre de critères. De nombreux articles portent sur la mesure de la centralité des agents. Il existe différentes définitions de la centralité qui renvoient toutes à des indicateurs différents (Bonacich, 1972 ; Freeman, 1979...). Ces derniers sont adaptables en fonction des différents réseaux que l'on étudie.

Grâce à ces indicateurs, il est possible d'identifier les "hubs" des réseaux, c'est-à-dire les agents dont le nombre de liens est important (Malerba, Vornortas, Breschi et Cassi, 2006 ; Roediger-Schluga et Barber, 2008 ; Deroïan, M'Chirgui et Milleli, 2007). Ces "hubs" ont un statut particulier au sein des réseaux dans le sens où il est nécessaire pour un réseau d'en posséder car ces agents au travers de leurs relations sont des plaques tournantes du réseau et permettent de réduire les distances entre les agents et améliorent l'efficacité du réseau. Il peut donc être nécessaire de disposer de tels agents. Dans le même temps, ces agents concentrent un grand nombre de liens qui leur confère un statut important qui rend le réseau dépendant des "hubs" d'une période à l'autre. Il est donc nécessaire d'avoir un nombre suffisant de "hubs" pour assurer l'efficacité du réseau et ne pas être dépendant d'un nombre réduit d'agents.

Théorie de la formation des réseaux et déterminants des choix de collaborations

De nombreux travaux aussi bien empiriques que théoriques sont venus compléter ces analyses en s'intéressant à la formation du réseau et plus précisément aux déterminants qui incitent les agents à collaborer en matière de R&D.

L'article de C. D'Aspremont et A. Jacquemin (1988) a ouvert la voie à de tels travaux. Les auteurs ont ainsi montré dans le cas d'un duopole que les

firmes avaient intérêt à collaborer. Le modèle est construit en deux étapes. A la première étape, les firmes choisissent un niveau d'effort en R&D. Ce niveau impacte sur le coût de production. A la seconde étape, les firmes choisissent les quantités produites pour le marché. Au cours de ces deux étapes, les décisions peuvent se prendre de manière coordonnée afin de maximiser le profit joint. On parle dans ce cas de comportement de collusion. Ces décisions peuvent également être prises de manière non coordonnée. Dans ce cas, chaque firme prend la décision qui maximise son propre profit sans égard pour le profit de son concurrent. C. D'Aspremont et A. Jacquemin montrent alors qu'il est préférable pour les entreprises de coopérer en R&D que les firmes collaborent ou non à la seconde étape. Ils mettent en évidence le fait que l'intensité des externalités accroît l'intensité de la collaboration en R&D. Ce travail se focalise sur les situations de duopole. Or, de nombreux marchés sont en situation d'oligopoles (avec un nombre d'agents supérieur à deux). Dans ce cas, les collaborations bilatérales ont des effets d'externalités sur les firmes non impliquées dans la collaboration. D'où l'intérêt des travaux prolongeant le travail de C. D'Aspremont et A. Jacquemin. Il s'agit dans ce cas de s'intéresser à la formation de coopérations de manière plus globale en cherchant à modéliser la formation de réseau au sein d'un ensemble d'agents.

Il existe en fait différentes approches sur la formation des réseaux en fonction du cadre choisi. Une distinction importante existe entre liens coopératifs et liens non coopératifs. Dans le premier cas, les agents joignent leur effort dans un but commun ou parce qu'il est mutuellement bénéfique pour les agents de collaborer. Les agents acceptent donc mutuellement de collaborer l'un avec l'autre. Dans ce cas, les agents sont souvent liés par des contrats mais ce n'est pas une obligation. Les raisons qui incitent deux agents à collaborer peuvent être diverses (partage des coûts, compétences complémentaires, atteinte d'une masse critique, bénéfice mutuel pour l'image, partage des moyens de production, échange de chercheurs,...). Dans le second cas, la décision de former un lien avec un autre agent ne requiert pas l'accord des deux parties (Bala et Goyal, 2000). Ce cadre permet ainsi de modéliser des situations d'intelligence économique dans lesquelles un agent peut créer des liens à l'insu des autres, par exemple des cas d'espionnage industriel (Billand et Bravard, 2004). Dans le cas des PCRD, les liens coopératifs sont la forme la plus à même de modéliser les situations d'interactions au sein des réseaux.

Concernant les liens coopératifs, Goyal et Joshi dans leur article "Unequal Connections" (2006) proposent une modélisation capable de couvrir un grand nombre de situations en termes de formation de réseau parmi lesquelles les réseaux de collaborations inter-firmes. En effet, ils présentent un modèle dans lequel la fonction de profit des agents est relativement générale. A partir des fonctions de profit individuel, ils construisent des fonctions de gain marginal

individuel qui dépendent de deux arguments : d'une part, du nombre de liens de l'agent qui forme le lien et, d'autre part, du nombre de liens du partenaire potentiel et/ou de l'ensemble des agents dans le réseau. A partir de là, ils modélisent deux situations bien distinctes selon l'hypothèse que l'on retient sur les externalités. Il crée ainsi deux jeux distincts. Dans le premier jeu, "Playing the field game", la fonction de gain marginal d'un agent suite à la création d'un lien dépend du nombre de liens de cet agent et de l'ensemble des liens existants dans le réseau avant la création du lien. Ce type de jeu a été examiné par Goyal et Joshi (2003). Dans le second type de jeu, "Local spillovers game", la fonction de gain marginal du joueur suite à la création d'un lien dépend du nombre de liens de l'agent et du nombre de liens de son partenaire potentiel avant la création du lien (Jackson et Wolinsky, 1996). De plus, Goyal et Joshi supposent que l'agent peut supprimer un ou plusieurs liens simultanément alors que la modélisation la plus courante ne permet de supprimer qu'un seul lien à la fois. Les modèles ainsi constitués permettent de modéliser différentes situations. Goyal et Joshi (2006) montrent ainsi, qu'en fonction des hypothèses retenues sur les externalités, la forme du réseau d'équilibre change (la notion d'équilibre retenue est le "pairwise stable equilibrium"⁶). Ces modèles permettent de mettre en avant certaines formes de réseaux d'équilibre comme le réseau complet (tous les agents sont connectés), ou le réseau en forme d'étoile (un agent est connecté à tous les autres, ces derniers ne sont connectés qu'à l'agent central),... Ils permettent aussi bien de modéliser des situations d'oligopole (Goyal et Joshi, 2003) que des réseaux de co-auteurs (Jackson et Wolinsky, 1996 ; Goyal, van der Leij et Moraga-Gonzalez, 2006) ou de co-brevets. Ces modèles apparaissent pertinents pour modéliser des situations proches de celles des PCRDT. Toutefois, il manque un élément essentiel pour pouvoir couvrir le cas des PCRDT : la diversité des agents. Goyal, van der Leij et Moraga-Gonzalez (2006) introduisent cette diversité dans un modèle de co-auteurs en supposant qu'il existe deux types d'agents : des agents aux idées d'une qualité élevée et des agents dont les idées sont d'un niveau plus faibles. Les idées des agents de niveau élevé sont donc plus à même d'augmenter le profit des agents. Les auteurs montrent alors qu'un réseau en forme d'étoiles inter-reliées à l'équilibre est possible. Or, c'est la forme de réseau qu'ils avaient identifiés dans leur étude empirique préalable à la construction de leur modèle théorique.

Il est possible de faire un certain nombre d'hypothèses susceptibles de modifier le réseau d'équilibre, notamment sur la manière dont les agents peuvent bénéficier des collaborations des autres joueurs dans le réseau. Jack-

⁶Pour un article sur les différentes formes d'équilibre dans les jeux de formation de réseau voir Bloch et Jackson, 2006.

son et Wolinsky (1996) utilisent l'hypothèse de réseau borné (ou tronqué) qui suppose qu'un agent ne peut pas bénéficier des collaborations des autres agents si ceux-ci sont situés à une distance⁷ supérieure à un certain seuil. Une autre hypothèse qu'il est possible d'utiliser est celle de "decay" qui renvoie au fait que plus la distance entre deux agents est élevée moins ceux-ci bénéficient mutuellement de ces travaux. Chaque lien supplémentaire conduit à une perte d'information et de connaissance. Il est aussi possible de faire varier les coûts de formation des liens (Goyal et Joshi, 2003). Ces différentes hypothèses permettent de se rapprocher un peu plus de la réalité et notamment d'un réseau comme celui des PCRDT. En effet, il est intuitif de supposer que les coûts supportés par les agents vont faire varier leur nombre de collaboration ainsi que le choix de leurs collaborateurs. De la même manière, leur implication au sein du réseau sera différente en fonction de la manière dont les connaissances, les informations et les externalités circulent au sein du réseau. La plupart des modèles théoriques n'ont pas donné lieu à des vérifications empiriques. Parmi les articles cités dans les paragraphes précédents, seul l'article de Goyal, van der Leij et Moraga-Gonzalez (2006) réalise une étude empirique. Toutefois, il faut noter que dans le cas de cet article, les auteurs commencent par réaliser leur étude empirique et dans un second temps construisent un modèle susceptible de produire des résultats proches de ceux constatés. Si peu de travaux réalisent des vérifications empiriques, on peut noter qu'une partie de la littérature cherche à montrer que les réseaux théoriques possèdent des propriétés similaires aux réseaux réels. Carayol et Roux (2008, 2009) utilisent le modèle de Jackson et Wolinski (1996) (une variation de ce modèle) pour mettre en avant le fait que certains réseaux théoriques, sous certaines conditions et certaines hypothèses, possèdent des propriétés proches des réseaux réels. De plus, Carayol et Roux (2009) montrent que les réseaux théoriques peuvent posséder des propriétés similaires aux "small-world networks".

Cependant, le cadre théorique des réseaux coopératifs ne correspond qu'imparfaitement à la réalité des PCRDT. En effet, dans le cas des PCRDT, les agents sont regroupés au sein de projets. Lorsqu'un agent accepte de prendre part à un projet, il accepte de collaborer avec l'ensemble des agents du projet. On peut alors supposer que le sous-ensemble d'agents impliqués dans le projet acceptent simultanément de fournir des liens entre eux. Ceci laisse penser qu'une modélisation des PCRDT en termes de formation de coalitions est plus appropriée. Toutefois, les PCRDT ne correspondent pas non

⁷On parle ici de distance en termes de nombre de liens séparant les agents. Nous reviendrons sur ce concept dans partie 1 de la thèse.

plus à ce cadre puisque dans les modèles basés sur les coalitions, les agents ne peuvent appartenir qu'à une seule coalition. Or, dans les PCRDT, un agent peut appartenir à plusieurs projets simultanément.

Au delà de la définition des réseaux d'équilibre, la littérature sur la formation des réseaux dans un cadre coopératif apporte un certain nombre d'éléments et d'hypothèses qu'il est important de connaître lorsque l'on étudie les réseaux réels. Les hypothèses testées par ces modèles ("decay" et réseau borné) sont proches de ce que l'on peut supposer dans la réalité. De plus, la lecture des différents articles théoriques permet de faire le constat que les collaborations entre agents sont déterminées par un grand nombre de facteurs et d'enjeux qu'il n'est pas simple d'analyser. Les décisions des agents sont influencées par leurs objectifs, leurs caractéristiques propres (Goyal, van der Leij et Moraga-Gonzalez, 2006) mais aussi par les décisions des autres agents. Dans le cadre de la thèse, qui est un travail empirique sur les réseaux au sein des PCRDT, il sera donc important de tenir compte de ces différents points notamment lorsque l'on étudiera le rôle joué par les agents au sein des PCRDT ainsi que leurs choix de collaborations.

De nombreux travaux sont venus compléter ces analyses en s'intéressant à la formation du réseau et notamment aux déterminants qui poussent les agents à collaborer (Fafchamps, van der Leij et Goyal, 2010 ; Autant-Bernard et al., 2007). Pour cela, les économistes recourent de plus en plus à l'économétrie et développent des méthodes pour construire les variables spécifiquement pour l'étude des réseaux (Fafchamps, van der Leij et Goyal, 2010). Il existe deux types d'approches : celles qui cherchent à expliquer le fait qu'un agent collabore ou non quel que soit ses partenaires (Cassiman et Veugelers, 2002) et celles qui cherchent à expliquer la formation de liens bilatéraux (Fafchamps, van der Leij et Goyal, 2010 ; Autant-Bernard et al., 2007 ; Paier et Scherngell, 2008). Le type d'approche que nous retenons dans la thèse est la seconde puisque nous cherchons à expliquer la formation du réseau. Les déterminants que les économistes cherchent à tester sont issus de l'économie de la connaissance et les travaux sur la formation de réseaux. Ainsi, parmi les déterminants traditionnels, on trouve les moyens en R&D qui peuvent renvoyer à différents indicateurs comme les dépenses en R&D, le nombre de chercheurs, la proximité institutionnelle, la proximité sectorielle, la proximité technologique,...

Dans les articles cherchant à expliquer la formation des liens, deux autres déterminants apparaissent particulièrement importants : la proximité géographique et la proximité sociale. La proximité sociale renvoie aussi bien au concept de distance au sein des réseaux qu'au fait que deux agents aient déjà collaboré ou possèdent un partenaire commun. Comme le mentionne Ozman (2009, p53) : *"Previous empirical research has confirmed that face-*

to-face contacts and geographical proximity are important factors facilitating the diffusion of innovations (Jaffe, Trajtenberg, and Henderson 1993), fostering some forms of knowledge exchange (Morgan 2004; Gomes-Casseres, Hagedoorn, and Jaffe 2006) and provide better access to information (Porter 1990).” De plus, Barabasi et Albert (2002) ont ainsi mis en avant la notion de “preferential attachment” qui veut qu’un agent attractif le soit de plus en plus au fil des réseaux créant ainsi un déséquilibre de plus en plus grand entre les agents au sein de réseaux en termes de nombre de liens. De manière différente, au travers d’interview de responsables au sein des entreprises, Gulati (1993, 1998) a mis en avant le fait que les agents pouvaient prendre en compte l’avis de leur partenaires pour déterminer leurs futures collaborations mettant ainsi en avant que les collaborations passées peuvent influencer les choix de collaborations futures.

La proximité géographique est supposée avoir plusieurs effets bénéfiques sur les performances des agents : premièrement, elle permet de réaliser des gains d’efficacité dans la collaboration (Colletis et Winterhalter, 1991 ; Graitson, 2000) et, deuxièmement, elle permet de réduire les coûts (Englmann et Walz, 1995 ; Martin et Ottaviano, 1999 ; Baranes et Tropeano, 2003). Ces différents travaux montrent bien le rôle important joué par la proximité géographique. Dans le cas des PCRDT, la proximité géographique est un déterminant un peu particulier dans la mesure où comme nous l’avons écrit l’UE a mis en place des incitations pour favoriser les collaborations entre agents de pays différents. Les premiers travaux d’Autant-Bernard et al. (2007) montrent que la proximité géographique joue un rôle ambigu dans la mesure où sur l’ensemble des collaborations dans le domaine IST (Information Society Technology) dans le 6^{ème} PCRDT, la proximité géographique ne joue pas un rôle déterminant. Par contre, lorsqu’ils s’intéressent seulement aux collaborations entre firmes françaises, Autant-Bernard et al. montrent que la proximité géographique a un effet positif. Ce phénomène s’explique alors en partie par l’introduction dans le même modèle à la fois de la proximité géographique et de la proximité sociale⁸. L’effet de la proximité sociale prenant alors le pas sur l’autre. Fafchamps, van der Leij et Goyal (2010) ont montré que les collaborations antérieures entre les agents modifiaient par la suite les choix de collaborations.

Les travaux économétriques sur la formation des réseaux sociaux sont de plus en plus nombreux. Les économistes cherchent à adapter les modèles ainsi que les variables à ce cadre particulier. Ces nouveaux développements

⁸La proximité sociale entre deux agents renvoie au fait que deux agents puissent soit avoir déjà collaboré dans le passé, soit posséder des partenaires communs ou soit être situé à proximité (en termes de nombre de liens) au sein du réseau social.

tendent à montrer que l'économétrie fait partie intégrante des outils propres à l'analyse des réseaux sociaux.

Apports de la thèse

Le travail réalisé dans la thèse met en avant les apports suivants :

- **Sélection d'indicateurs pertinents et création de nouveaux indicateurs.** Une sélection des indicateurs jugés les plus pertinents afin d'étudier les réseaux de collaborations se déroulant dans le cadre des PCRDT est opérée. L'analyse des réseaux sociaux compte un nombre important d'indicateurs susceptibles d'étudier les réseaux de collaborations. Cependant, certains apparaissent plus pertinents en fonction de l'objet étudié et de l'analyse réalisée. Dans le cadre de la thèse, nous mobilisons les indicateurs les plus pertinents pour étudier la structure des réseaux de collaboration des PCRDT ainsi que le rôle des agents au sein des PCRDT. Les indicateurs présentés proviennent des travaux menés par Bonacich (1978, 1987, 1991), Freeman (1979), Faust (1997), Borgatti et Everett (1997, 1999, 2000),... De plus, ce travail est complété par la mise en place de nouveaux indicateurs destinés à contourner certaines difficultés rencontrées par les indicateurs usuels dans certaines circonstances.
- **Une analyse à deux niveaux.** L'analyse menée dans cette thèse se situe au niveau individuel puisque nous construirons des réseaux de collaborations entre agents mais aussi à un niveau plus agrégé puisque nous créons des réseaux entre régions à partir des réseaux individuels. Il est ainsi possible d'obtenir des informations complémentaires sur la structuration des réseaux, informations qui permettent de vérifier si les différents objectifs poursuivis par les PCRDT ont été vérifiés. L'analyse des réseaux au niveau individuel est plus courante (Breschi et Cusmano, 2002 ; Wagner, Thomson, tesch et Perez, 2004 ; Rossi, 2005 ; Malerba, Vonortas, Breschi et Cassi, 2006 ; Roediger-Schluga et Barber 2008). Elle permet à la fois d'observer la structure des réseaux de collaborations ainsi que le rôle des agents au sein des réseaux. Les réseaux de régions permettent de connaître les régions qui possèdent le plus d'agents et qui participent aux PCRDT et surtout de mettre en avant les axes de collaborations privilégiés au sein de l'UE dans les PCRDT. A partir de ces réseaux, il est alors possible de réaliser un comparatif entre les régions leaders dans les PCRDT et les régions leaders de manière générale dans l'innovation. Enfin, les réseaux de régions permettent d'observer si les nouveaux États membres occupent

une place plus importante au fil des PCRDT et dans quelle mesure l'objectif de cohésion est atteint.

- **L'estimation du rôle joué par la proximité sociale.** Dans la littérature qui s'intéresse aux incitations des agents à collaborer, il existe un certain nombre de déterminants qui sont couramment testés. Deux des principaux déterminants sont les moyens en R&D et la proximité géographique. Dans la thèse, nous introduisons la proximité sociale. Nous chercherons ainsi à montrer que la probabilité que deux agents collaborent est plus élevée si les deux agents ont déjà collaboré ensemble ou s'ils possèdent une connaissance commune. De cette manière, il est possible d'observer dans quelle mesure les collaborations antérieures affectent les collaborations présentes. Pour compléter notre analyse, nous examinons aussi dans quelle mesure le rôle joué par un agent dans le réseau de la période antérieure (mesuré par un indicateur de centralité) influe sur ces choix de collaborations dans le réseau présent. Plus particulièrement, nous cherchons à vérifier si les agents centraux ont tendance à collaborer entre eux.
- **L'analyse de l'existence et de l'intensité des collaborations.** La plupart des études qui portent sur les PCRDT s'intéressent uniquement à l'existence ou non d'une collaboration entre deux agents. On parle alors de réseaux binaires (Breschi et Cusmano, 2002 ; Wagner, Thomson, Tesch et Perez, 2004). Or, une information importante que cache ses représentations est la possibilité que deux agents collaborent plus d'une fois ensemble dans le même programme cadre. La prise en compte de l'intensité des liens conduit à la construction de réseaux valués. L'intensité des liens entre les agents peut permettre d'affiner l'analyse des PCRDT et notamment faire apparaître des cliques intensément connectées (Roediger-Schluga et Barber, 2008). Au delà des cliques, l'intensité des liens est très importante à prendre en compte lorsque l'on étudie le rôle des agents au sein des réseaux. Certains agents peuvent posséder un faible nombre de partenaires mais entretenir des collaborations variées avec ces partenaires soit en collaborant sur des projets différents dans un même domaine technologique soit en collaborant dans des domaines différents mais complémentaires. De plus, on peut supposer que les déterminants des collaborations entre deux agents sont différents selon que l'on s'intéresse à la formation du lien ou à son intensité. Dans le cadre des travaux économétriques menés dans cette thèse, nous chercherons à vérifier si les déterminants des collaborations qui expliquent la formation des liens peuvent aussi permettre d'expliquer l'intensité de ces derniers. Enfin, lorsque l'on étudie les réseaux de régions, nous montrerons que la valeur des liens entre les régions prend un sens diffé-

rent en fonction du choix initial adopté pour construire le réseau entre agents.

- **Un cadre d’analyse prolongeant le cadre des réseaux de relations bilatérales avec la prise en compte des triades.** De manière générale, lorsque l’on s’intéresse aux incitations des agents à collaborer, nous étudions la formation de liens entre deux agents. Dans, le cadre des PCRDT, il peut apparaître complémentaire d’étudier la formation de sous-ensemble de trois agents. En effet, les PCRDT possèdent un cadre particulier qui implique notamment que nombre de projets doivent contenir au moins trois agents. Il apparaît alors que l’étude de la formation de liens bilatéraux entre agents est limitée dans la mesure où une partie des stratégies d’incitation est guidée par le fait que les agents ne doivent pas choisir un partenaire mais au moins deux. L’analyse par les triades apparaît alors comme un complément idéal de l’étude de la formation des liens bilatéraux. Ces deux analyses, liens bilatéraux et triades, apparaissent complémentaires dans la mesure où lorsqu’un agent doit choisir ses collaborateurs, il accepte à la fois de travailler avec chacun d’eux (collaborations deux à deux) et, dans le même temps, il est contraint par les conditions imposées par l’UE (triade).

Structuration de la thèse

Afin de rendre compte de ces apports, la structure de la thèse est organisée en deux parties.

La première partie cherche à étudier les réseaux formés par l’ensemble des collaborations réalisées dans deux domaines technologiques du 6^{ème} PCRDT appartenant à la thématique Information Society Technology (IST). Dans le premier chapitre, nous présentons les différentes représentations possibles pour les réseaux de collaborations et nous justifions le choix fait dans la thèse. Puis, nous présentons l’ensemble des indicateurs retenus pour mener à bien l’analyse ainsi que des indicateurs supplémentaires créés afin de contourner certaines difficultés rencontrées. Dans le second chapitre, nous étudions la structure des réseaux et le rôle joué par les agents. Nous cherchons à vérifier si les réseaux de collaborations au sein des PCRDT possèdent des propriétés favorables à l’innovation. Le fait de retenir deux domaines technologiques différents n’est pas anodin dans la mesure où nous montrerons qu’un des qualités des réseaux des PCRDT est l’interconnexion qu’il existe entre les domaines technologiques. Dans l’optique de vérifier si les réseaux de collaborations des PCRDT possèdent des propriétés favorables à la diffusion des connaissances et à l’innovation, nous montrerons l’évolution de la structure

du réseau entre le 5^{ème} PCRDT et le 6^{ème} PCRDT. Ensuite, nous nous intéresserons aux rôles joués par les agents dans ces réseaux et notamment à la place prise par les agents les plus centraux. Enfin, à partir des réseaux étudiés jusqu'alors, nous construirons des réseaux agrégés au niveau des régions afin de nous intéresser à la manière dont se répartissent les agents et les collaborations sur le territoire européen dans le cadre des PCRDT ainsi qu'à la manière dont s'intègre les nouveaux États membres.

Dans la seconde partie de la thèse, nous chercherons à montrer quelles sont les déterminants des choix de collaborations des agents qui ont conduit à la formation des réseaux de PCRDT, déterminants qui ont conditionné la structure des réseaux étudiés au préalable. Nous nous intéresserons aux facteurs qui permettent d'expliquer, au moins en partie, l'existence d'un lien entre deux agents. Pour cela, nous chercherons à vérifier si les déterminants traditionnels de la littérature sur les collaborations (moyens mis en œuvre par les agents dans les activités R&D, proximité géographique, ...) sont vérifiés dans le cadre des PCRDT. En plus des déterminants traditionnels, nous chercherons à montrer le rôle joué par la proximité sociale, plus précisément les collaborations s'étant déroulées dans le 5^{ème} PCRDT, affectent les choix des collaborations dans le 6^{ème} PCRDT. Après avoir identifié les principaux facteurs de l'existence d'une collaboration entre deux agents, nous réaliserons un certain nombre d'estimations afin de vérifier si les déterminants qui expliquent l'existence d'un lien entre deux agents permettent aussi d'expliquer l'intensité des liens entre deux agents. Enfin, dans le dernier chapitre de la thèse, nous introduirons à un nouveau cadre d'analyse en étudiant la formation de triades au sein des PCRDT. Nous adoptons cette approche afin d'observer notamment si les contraintes imposées par l'UE pèsent sur les choix de collaborations et surtout afin de compléter notre analyse sur le rôle joué par les collaborations à la période antérieure puisque nous essaierons de vérifier si au sein des PCRDT on peut identifier un phénomène de "triadic closure".

Première partie

Structure et positionnement au
sein des réseaux de collaborations
en R&D appliqués aux PCRDT

Introduction

L'objectif de cette partie est d'étudier les propriétés des réseaux formés par les collaborations au sein des Programmes Cadres de Recherche et Développement Technologique (PCRDT). Cette étude se fera à plusieurs niveaux. Tout d'abord, à un niveau macroéconomique, nous étudierons les propriétés générales du réseau. Nous chercherons à démontrer si les réseaux formés par les PCRDT possèdent des propriétés favorables à l'innovation. Puis, à un niveau microéconomique, nous étudierons la place occupée par certains agents au centre de ces réseaux de collaborations. Nous montrerons que les réseaux possèdent un certain nombre d'agents centraux qui sont des rouages essentiels pour leur bon fonctionnement. Enfin, à un niveau régional, nous mettrons en avant la prépondérance de certaines régions dans les collaborations au sein des PCRDT. Certaines régions se distinguent par des collaborations importantes dans les différents domaines technologiques étudiés, d'autres apparaissent comme des régions plus spécialisées dans un des deux domaines technologiques. Lorsque l'analyse réseau le permettra, nous chercherons à examiner dans quelle mesure les objectifs fixés aux PCRDT ont été réalisés. Nous montrerons alors dans quelle mesure l'analyse réseau est utile pour une telle évaluation.

Nous nous intéressons principalement au 6^{ème} PCRDT⁹ car c'est le dernier PCRDT pour lequel nous disposons de données complètes. Ce programme cadre est le premier qui fait suite au sommet de Lisbonne en 2000, sommet qui met en avant la volonté de l'UE de construire un Espace Européen de la Recherche (EER). De plus, il se déroule dans un contexte d'adhésion de nouveaux États à l'Union Européenne. L'objectif principal du 6^{ème} PCRDT est : *“to contribute to the creation of the European Research Area (ERA) by improving integration and coordination of research in Europe which is so far largely fragmented”* (Source : Europa¹⁰).

⁹Une présentation plus détaillée du 6^{ème} PCRDT, qui précise notamment les différentes thématiques, types d'agents et instruments, est jointe en annexe A.1

¹⁰Site internet : www.europa.eu.

Nous nous intéresserons seulement aux projets de collaborations qui se déroulent dans le cadre des 7 priorités thématiques retenues par l'Union Européenne. Ce choix s'explique par le fait que l'essentiel des moyens mis en œuvre par l'UE dans le cadre des PCRDT est utilisé pour soutenir les projets collaboratifs qui se déroulent dans ces priorités thématiques. En ne prenant en compte que les collaborations qui se déroulent dans les 7 priorités thématiques, on limite le biais que pourrait entraîner la prise en compte de dispositifs incitatifs spécifiques. Par exemple, en dehors de ces 7 priorités thématiques, il existe un dispositif qui favorise les collaborations entre PME (projets de recherche coopératif CRAFT ou projets de recherche collectif). Prendre en compte les collaborations qui se déroulent dans ce dispositif entraînerait une sur-représentation des PME dans les collaborations et rendrait difficile une étude des déterminants des choix de collaboration (cf. partie 2) du fait que les incitations à collaborer au sein des thématiques sont différentes de celles qui gouvernent le choix de participer à ce programme spécifique.

Il est difficile d'étudier l'ensemble des collaborations présentes dans les 7 priorités thématiques. Pour notre part, nous nous intéresserons seulement à deux domaines technologiques de la priorité thématique "Information Society Technology" (IST) : le domaine des Télécommunications et le domaine de l'Electronique-Microélectronique. Cette priorité est celle qui capte la plus grande partie des financements. En effet, 3,95 milliards d'euros sont alloués à la thématique IST sur les 12,4 alloués aux 7 priorités thématiques (soit 32% du budget alloué aux thématiques). Les domaines Télécommunications et l'Électronique-Microélectronique sont les domaines dans lesquels on recense le plus de participations. Retenir deux domaines différents autorisent des comparaisons entre leurs réseaux respectifs et permet d'examiner l'existence de connections entre des domaines différents. Cette restriction implique cependant que les conclusions en termes de réalisations des objectifs des PCRDT ne seront que partielles et valables uniquement sur les domaines étudiés. Les apports méthodologiques sont eux applicables à tout programme cadre d'incitation à coopération par financement de projets collaboratifs.

L'ensemble des informations dont on dispose pour construire les réseaux et mener notre étude empirique est présenté en annexe A.2. La principale difficulté réside dans l'identification de l'ensemble des agents qui participent aux 6^{ème} PCRDT. Pour pouvoir construire les réseaux de collaborations, les deux informations obligatoires sont l'ensemble des projets et l'ensemble des agents qui participent à chacun des projets. Au sein des programmes cadres, de nombreux agents participent à plusieurs projets. Il est primordial de pouvoir identifier clairement ces agents car ce sont eux qui, en reliant les projets, définissent le niveau de connexion globale des réseaux. Cette identification est un processus long et difficile d'autant plus que pour pouvoir évaluer de

manière plus précise les PCRDT et réaliser une étude économétrique (partie II), il est nécessaire de posséder plusieurs types d'informations sur chaque agent au delà de son simple nom¹¹.

La partie 1 suit le plan suivant. Le chapitre 1 est consacré à la présentation des outils de l'analyse réseau qui semblent les plus pertinents pour caractériser les réseaux de collaborations des PCRDT et évaluer dans quelle mesure les programmes cadres remplissent leurs objectifs. Après avoir présenté rapidement certaines notions de base (section 1.1), nous présenterons les différentes représentations possibles. Parmi ces représentations, nous en choisissons une et nous justifions ce choix (section 1.2.). Par la suite, nous présentons les différents indicateurs sélectionnés pour étudier les réseaux de collaborations aussi bien au niveau macroéconomique (section 1.3.) qu'au niveau microéconomique (section 1.4). Enfin, nous présentons l'approche cœur/périphérie qui apporte un complément d'analyse aux indicateurs précédemment retenus (section 1.5).

Les PCRDT sont un instrument de politique économique. Il est important de montrer que l'analyse des réseaux permet d'apporter, au travers des différentes analyses que l'on pourra mener, un certain nombre d'éléments pour évaluer les PCRDT et qu'elle constitue en ce sens un outil complémentaire aux évaluations menées couramment. Les PCRDT ont déjà fait l'objet de plusieurs études et rapports qui visent à vérifier si les objectifs fixés aux PCRDT ont été atteints. Nous pouvons notamment citer les articles de Breschi et Cusmano (2004) et de Malerba, Vonortas, Breschi. et Cassi (2006). Le premier article porte sur l'ensemble des projets financés dans le cadre des 3^{ème} et 4^{ème} PCRDT. Le second utilise une approche différente car ils n'utilisent pas seulement les données propres aux PCRDT mais aussi des données sur des collaborations qui portent sur des contrats de R&D hors PCRDT et sur des citations de brevets. Les auteurs retiennent cette approche afin de voir si les agents les plus centraux de ces différents réseaux participent aux PCRDT et, inversement, si les agents centraux des PCRDT occupent une position centrale dans les autres réseaux.

Dans le chapitre 2, nous mobilisons les outils présentés dans le chapitre 1 pour étudier plusieurs dimensions des réseaux de PCRDT et montrer quel peut être l'apport de l'analyse réseau pour mieux appréhender ces collaborations et le réseau qu'elles forment. Nous analyserons les réseaux de collaborations entre agents (section 2.1) afin d'identifier les propriétés qui les caractérisent. Il sera intéressant par la suite de vérifier si ces propriétés ap-

¹¹La méthode utilisée pour identifier les agents participants aux PCRDT est présentée dans l'annexe A.3.

paraissent comme favorable à l'innovation. Puis, dans la section 2, nous mettrons en avant que les agents occupent des positions différentes et jouent un rôle différent. En effet, un certain nombre d'agents concentrent une grande partie des liens et apparaissent comme des agents centraux. Au contraire, certains agents participent peu et occupent une position plus périphérique. La centralité renvoie à la position qu'occupe un agent au sein du réseau. Il existe différentes notions de centralité en fonction de ce que l'on souhaite mesurer. Enfin, nous montrerons qu'il est utile d'étudier les réseaux de relations entre régions obtenus à partir des collaborations réalisées dans le cadre des PCRDT afin de compléter l'analyse réalisée à un niveau individuel et de pouvoir ainsi évaluer de manière différente et complémentaire la politique des PCRDT (section 2.3).

Chapitre 1

Définitions et apports statistiques de l'analyse des réseaux.

Ce chapitre présente un certain nombre d'indicateurs issus de l'analyse des réseaux sociaux. Il n'a pas vocation à décrire l'ensemble des indicateurs réseaux existants mais ceux qui nous semblent pertinents dans la poursuite de notre objectif qui est d'évaluer la politique des PCRDT.

Dans la section 1.1, nous rappelons rapidement quelques notions de base de l'analyse des réseaux sociaux qui nous seront utiles par la suite. La section 1.2 met en avant de quelle manière, à partir des données dont nous disposons, nous construisons les réseaux de collaborations et quelles sont les représentations possibles pour ces réseaux. La section 1.3 présente les indicateurs de structure nécessaires à l'étude des réseaux pris dans leur globalité. Ces indicateurs sont notamment utiles pour comparer les réseaux des deux domaines technologiques et déterminer si ces réseaux sont propices à l'innovation. La section 1.4. présente un ensemble d'indicateurs de centralité pour pouvoir étudier la position occupée par les agents au sein des réseaux. Ces indicateurs sont utiles pour déterminer quels sont les agents importants dans le fonctionnement du réseau. Enfin, la section 1.5 décrit une méthode appelée cœur/périphérie qui vise à trouver une partition de l'ensemble des agents en deux groupes. Chaque groupe ayant ses propres caractéristiques. Cette méthode est utile pour déterminer un sous-ensemble d'agents qui se distinguent par de nombreuses collaborations et qui constituent l'ossature du réseau.

Les indicateurs présentés dans ce chapitre seront mobilisés dans le chapitre suivant pour montrer l'apport de l'analyse des réseaux sociaux dans l'étude des PCRDT, plus particulièrement en termes d'évaluation des PCRDT. De plus, les indicateurs de centralité seront utilisés aussi dans la partie économétrique pour la construction de variables explicatives supplémentaires (cf. partie II).

1.1 Notions de bases

De manière générale, les réseaux sont représentés à l'aide de graphes composés d'un ensemble de points (d'agents dans notre cas) N et d'un ensemble de liens L . A chaque réseau est associé un graphe $G(N, L)$. L'ensemble des graphes est noté \mathcal{G} . Le graphe $G'(N', L')$ est le sous graphe de $G(N, L)$ induit par N' , avec $N' \subset N$ et $L' = \{g_{ij} = 1\} \in L \mid i \in N', j \in N'\}$ avec $g_{ij} = 1$ qui signifie que les agents i et j ont formé un lien, $g_{ij} = 0$ sinon.

Trois graphes sont particulièrement utilisés dans l'analyse réseau : le graphe vide, le graphe complet et le graphe en forme d'étoile. Le graphe vide est un graphe dans lequel il n'existe aucun lien entre les agents. A l'opposé, le graphe complet est un graphe dans lequel tous les liens potentiels entre les agents sont effectifs. Enfin, un graphe en forme d'étoile est un réseau dans lequel un agent est au centre de l'étoile et possède un lien avec tous les autres agents du réseau alors que ces derniers, dit agents périphériques, possèdent un seul lien, celui qui les relie à l'agent central.

Les liens présents au sein d'un réseau peuvent être :

- Orientés ou non orientés. Un lien est dit orienté si le sens du lien importe dans la relation par exemple dans le cas d'un don. Dans ce cas, $g_{ij} = 1$ n'est alors pas équivalent au lien $g_{ji} = 1$. Par opposition, un lien non orienté est un lien dans lequel le sens ne compte pas, c'est notamment le cas dans une collaboration. Dans ce cas, $g_{ij} = 1$ et $g_{ji} = 1$ sont équivalents ou autrement dit $g_{ij} = 1 \in L$ si et seulement si $g_{ji} = 1 \in L$. On parle alors de réseaux symétriques.
- Valués. Chaque lien peut renvoyer à une valeur. Il peut par exemple s'agir de flux d'argent ou d'un comptage, si au sein d'un réseau de collaborations on compte le nombre de collaborations entre deux agents. Le cas le plus courant est d'avoir recours au cas binaire dans lequel la valeur 1 sert à marquer la présence d'un lien entre deux points et 0 l'absence de liens.
- Directs ou indirects. On parle d'un lien direct entre deux agents lorsque ceux-ci possèdent un lien entre eux. Par exemple, entre un point i et un point j , si i et j ont un lien direct alors $g_{ij} \in L$. Il peut exister un lien indirect entre l'agent i et l'agent j s'il existe une succession de liens directs telle que $g_{i_0 i_1} = g_{i_1 i_2} = g_{i_2 i_3} = \dots = g_{i_{n-1} i_n} = 1$ avec $i_0 = I$, $i_n = j$ et un ensemble d'agents i_2, i_3, \dots, i_{n-1} .
- Des boucles (ou liens réflexifs). Une boucle est un lien dont le point de départ et le point d'arrivée sont identiques. Dans le cas des réseaux de collaborations, il est impossible d'avoir ce type de liens sur données individuelles.

Une succession d'agents et de liens commençant par un agent et finissant par un agent forme un chemin. Il existe différentes définitions d'un chemin utilisées dans l'analyse réseau. Pour notre part, nous définirons un chemin comme une succession d'agents et de liens tel que tous les agents et tous les liens sur le chemin sont distincts. Il est ainsi impossible de passer deux fois par le même agent ou de passer deux fois par le même lien. La distance d'un chemin est égal au nombre de liens qui le compose. Certains indicateurs utilisent un type de chemin particulier : le chemin géodésique. Le chemin géodésique est le plus court chemin pour relier deux agents. La distance géodésique est alors le nombre de liens sur ce chemin.

1.2 Des données de collaboration aux réseaux : le choix d'un mode de représentation

L'objectif de cette section est d'expliquer comment à partir de nos données, il est possible de construire les réseaux de collaborations qui seront ensuite l'objet de l'analyse. Il existe trois représentations possibles : les "two-mode networks" (sous-section 1.2.1), les "one-mode networks" (sous-section 1.2.2) et l'hypergraphe (sous-section 1.2.3) mais seulement deux sont compatibles avec une analyse réseau. Parmi ces deux représentations, nous n'en retiendrons qu'une seule pour poursuivre l'analyse. Nous montrerons que pour le type de représentation retenue, il existe deux sous-types de représentation qui permettent d'obtenir des éclairages complémentaires quant aux propriétés des réseaux de collaborations observés.

Pour construire les réseaux de collaborations propres aux PCRDT, deux informations sont nécessaires : les projets et le consortium (l'ensemble d'agent) propre à chaque projet. Le coordonnateur de chaque projet peut aussi être utile comme nous le verrons par la suite. A partir de ces informations il est possible de représenter les réseaux de trois manières différentes.

Pour illustrer les différentes représentations, nous nous appuierons à chaque fois sur l'exemple suivant (exemple 1) :

Projet	Identifiant
A	1*
A	2
A	3
B	1*
B	4
B	5
C	5*
C	6
D	3
D	7*
D	1

Exemple 1

Ce tableau décrit un ensemble de collaborations. Dans cet exemple, on dénombre 4 projets et 7 agents. Pour chaque projet, il y a un coordonnateur qui est indiqué par un astérisque (*). Par exemple, les agents 1, 2 et 3 participent au projet A et c'est l'agent 1 qui est coordonnateur de ce projet.

1.2.1 “Two-mode networks” : une première étape dans la construction

La première représentation possible n'est pas celle qui est la plus utilisée pour représenter les réseaux de collaborations des PCRDT. Il s'agit du “two-mode network” appelé aussi réseau d'affiliation ou bipartite graph. Sur le réseau, figurent à la fois les agents (les participants aux programmes cadres) et les projets. Les liens représentent simplement la participation d'un agent à un projet. Un projet est donc lié à tous les agents qui le composent et un agent est lié à tous les projets auxquels il participe. De ce fait, il n'existe aucun lien entre projets et aucun lien entre agents au sein des “two-mode networks”.

De manière formelle, ces réseaux sont composés de deux ensembles : un ensemble d'agents et un ensemble d'événements qui lient les agents entre eux. Dans le cas des PCRDT, les deux ensembles sont : les participants aux programmes cadres, noté N et les projets retenus par la Commission Européenne, noté M . Le nombre d'agents dans le réseau est n et le nombre de projets m . Les agents peuvent participer à plusieurs projets dans le même programme cadre. Les liens qui relient les agents aux projets appartiennent à l'ensemble L_A .

Nous pouvons représenter les affiliations entre les projets et les agents par une matrice que l'on note A . Un élément de A est noté a_{ib} avec $i \in N$ et avec $b \in M$. L'élément a_{ib} prend la valeur 1 si l'agent i participe au projet b et la valeur 0 sinon.

$$a_{ib} = \begin{cases} 1 & \text{si l'agent } i \text{ participe au projet } b \\ 0 & \text{sinon} \end{cases}$$

Chaque ligne de A donne la liste des projets auxquels participent un agent. La somme de la ligne donne le nombre de projets auxquels participe cet agent. Chaque colonne donne la liste des participants à un projet. La somme de chaque colonne donne le nombre de participants à un projet. Le nombre de liens du réseau est égal à :

$$l_A = |L_A| = \sum_{i=1}^n \sum_{b=1}^m a_{ib}$$

Afin d'illustrer la construction d'un "two-mode network", nous reprenons l'exemple 1. A partir du tableau présenté en début de section, nous obtenons la matrice A suivante qui décrit les relations projets-agents.

$$A = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Matrice d'affiliation de l'exemple 1

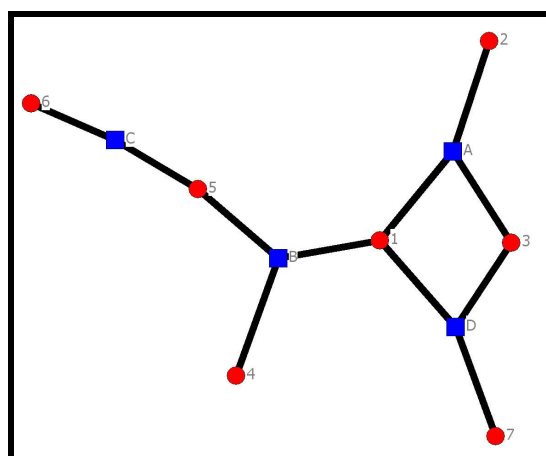
La matrice peut aussi être représentée sous la forme d'un tableau :

	Projet A	Projet B	Projet C	Projet D
Agent 1	1	1	0	1
Agent 2	1	0	0	0
Agent 3	1	0	0	1
Agent 4	0	1	0	0
Agent 5	0	1	1	0
Agent 6	0	0	1	0
Agent 7	0	0	0	1

Matrice d'affiliation de l'exemple 1

A partir de la matrice, il est alors possible de représenter le réseau sous la forme de graphes. Un graphe est composé d'un ensemble de points et d'un ensemble de liens qui relient les agents entre eux. Un graphe représentant un "two-mode network" est noté $g(N, M, L_A)$. Les graphes obtenus sont appelés des graphes bipartites.

A partir de l'exemple 1, nous obtenons le graphe suivant :



Graphe 1 : Représentation du réseau de l'exemple 1 en "two-mode network".

Sur le graphe, les projets et les agents apparaissent simultanément. Grâce à cette représentation, on connaît rapidement quels sont les agents qui composent un projet et quel est l'ensemble des projets auxquels un agent participe. Ce graphe permet de faire apparaître presque toute l'information relationnelle que l'on possède au sein de la base de données, seule l'information concernant le coordonnateur n'apparaît pas. De plus, on remarque que le lien qui relie un projet à un agent est non orienté, c'est-à-dire que le lien entre le projet b et l'agent i est équivalent au lien entre l'agent i et le projet b .

Appliquée au cas des PCRDT, la représentation des réseaux en "two-mode network" comporte plusieurs limites. Premièrement, contrairement aux agents, les projets appartiennent à un seul domaine technologique. Ce ne sont pas eux qui assurent la liaison entre deux domaines différents ou entre deux programmes cadres différents. Dans la mesure où dans notre étude des PCRDT, nous retenons deux domaines technologiques afin d'établir des comparaisons, il apparaît plus opportun de construire un réseau composé uniquement d'agents. Deuxièmement, représenter seulement les agents dans les réseaux de collaborations facilite la lecture de ces derniers. En effet, il est ainsi

possible de repérer plus directement les liens entre agents qui sont reconduits d'un programme à un autre ou encore qui sont présents dans deux domaines technologiques différents. Troisièmement, dans nos modèles économétriques, nous cherchons à identifier quels sont les déterminants des collaborations entre agents, nous ne nous intéressons pas aux relations agents-projets. La représentation en "one-mode network" présentée dans le paragraphe suivant permet de contourner ces limites et correspond mieux à l'objet de la thèse.

1.2.2 "One-mode networks" : des réseaux centrés sur les collaborations entre agents

Les "two-mode networks" permettent de représenter à la fois les projets et les participants des PCRDT sur les mêmes graphes. Il est également possible de représenter séparément les agents et les projets à l'aide des "one-mode networks". Ces réseaux se concentrent sur un seul ensemble à la fois et utilisent le second pour créer les liens entre agents. Ces réseaux correspondent à des graphes unipartites. Dans notre cas, nous souhaitons représenter uniquement les participants aux PCRDT sur le réseau, nous utiliserons alors la participation des agents aux projets pour créer des liens entre agents.

Il serait possible de la même manière de créer un réseau de projets reliés par la participation des agents. Toutefois, étant donnés les objectifs de la thèse, les informations complémentaires apportées par la représentation de réseaux de projets ne sont pas primordiales. En effet, nous souhaitons comparer les réseaux associés à chaque domaine technologique retenu. Or, les projets de collaborations ne sont déposés que dans un seul domaine technologique contrairement aux agents qui peuvent participer aux deux domaines technologiques et ne concernent qu'un seul programme cadre. De plus, nous souhaitons étudier dans la partie II les collaborations entre agents et, dans ce cadre, les relations entre projets apparaissent peu utiles. Ce sont les deux principales raisons qui font que par la suite nous n'utiliserons que des "one-mode networks" décrivant des relations entre agents.

Compte tenu des informations en notre possession, la représentation peut différer selon le rôle que l'on accorde au coordonnateur (paragraphe 1.2.2.1). De plus, il existe deux sous-types de représentation pour les "one-mode networks"(paragraphe 1.2.2.2.) :

- Pour prendre en compte l'intensité des liens, on utilisera des réseaux valués ;
- Pour étudier seulement l'existence des liens sans prendre en compte l'in-

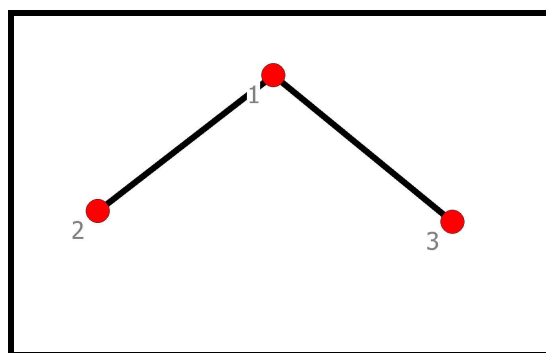
tensité, on utilisera alors des réseaux valués particuliers puisque ceux-ci seront binaires.

1.2.2.1. Hypothèses nécessaires à une représentation en “one-mode networks”

Pour construire les réseaux, nous avons vu que nous possédions comme informations les projets et le consortium associé à chaque projet. De plus chaque projet possède aussi un coordonnateur. Par contre, nous ne possédons pas d'informations plus précises sur les liens entre agents à l'intérieur du projet.

Dans les “two-mode networks”, le coordonnateur n'est pas pris en compte. Dans les “one-mode networks”, il est possible de prendre en compte le rôle spécifique joué par cet agent au sein du projet. Selon l'importance qu'on souhaite lui attribuer, il existe deux possibilités pour représenter les collaborations au sein d'un projet. Pour illustrer ces deux possibilités, reprenons le projet A de l'exemple 1. Ce projet est composé de 3 agents dont un coordonnateur l'agent 1.

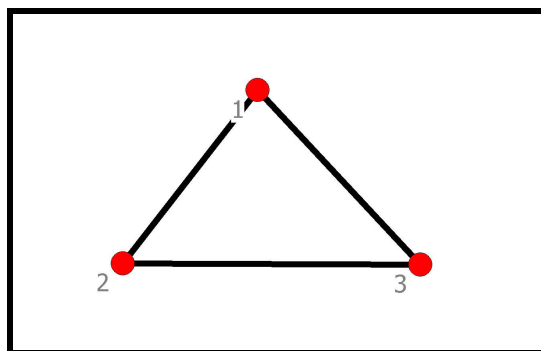
Le coordonnateur joue un rôle d'interface entre les agents du projet et la Commission européenne tout en étant chargé de centraliser les informations et de gérer l'organisation de la recherche au sein du projet. De ce fait, nous sommes certains qu'il est en relation avec tous les agents du projet tandis que nous ne possédons pas d'informations sur les relations entre les autres agents. A partir de ces éléments, nous pouvons considérer que le coordonnateur est l'agent central du projet et représenter les relations au sein du projet à l'aide d'un graphe en forme d'étoile dont le coordonnateur occupe le centre.



Graphe 2 : Le coordonnateur, un acteur central.

Dans ce cas de figure, il semble toutefois que le coordonnateur est trop mis en avant et que les collaborations au sein du réseau sont sous estimées.

De plus, cela revient à supposer que si deux agents participent au même projet et, qu'aucun des deux n'est coordonnateur, ils sont seulement reliés par le coordonnateur via un lien indirect. Cependant, s'il n'existe pas de collaborations directes et réelles entre deux agents d'un projet dans la réalité, on peut supposer qu'ils sont tout de même en plus étroite collaboration que deux agents qui ne participent pas au même projet. De plus, dans le cas où deux projets sont supportés par le même coordonnateur, il n'existe aucune différence entre les agents des deux projets. Or, on peut supposer que les connaissances et les externalités se diffusent plus facilement entre ces deux agents qu'entre des agents qui participent à deux projets différents. Afin de pouvoir construire un réseau de collaborations entre agents, nous devons faire une hypothèse forte à savoir que tous les agents qui collaborent dans un même projet ont un lien entre eux. Nous obtenons la représentation suivante.



Graphe 3 : Le projet comme réseau de collaboration.

Pour construire les “one-mode networks” valués et binaires, nous retenons la seconde hypothèse, c'est-à-dire que tous les agents collaborant au sein d'un même projet possèdent un lien entre eux, car c'est celle qui semble être la plus juste étant donné le fonctionnement des PCRDT, les informations dont on dispose (ou plutôt l'absence d'information plus précise sur le rôle joué par chaque agent dans les projets) et les objectifs que nous nous sommes fixés. De fait, il apparaît cohérent de lier l'ensemble de ces agents dans les réseaux de collaborations pour représenter le fait que la participation de ces agents à un même projet implique des passerelles entre ces agents dans l'échange de connaissance. Enfin, elle est cohérente avec les exigences de l'UE en matière de circulation des connaissances et d'informations au sein des projets. Notons que cette hypothèse est celle qui est traditionnellement retenue dans la littérature aussi bien dans les études statistiques (Breschi et Cusmano, 2004 ; Malerba et al., 2006 ; Fafchamps, van der Leij et Goyal, 2010 ; Breschi et Catalini, 2010 ; Deroian, M'Chirgui et Milleli, 2007 ; Roediger-Schluga et

Barber, 2008) qu'économétriques (Autant-Bernard et al., 2007 ; Arranz et Fernandez de Arroyabe, 2008 ; Hanaki, Nakajima et Ogura, 2007 ; Paier et Scherngell, 2008).

1.2.2.2. Représentation des “one-mode networks”

La collaboration entre deux agents est considérée comme symétrique d'où l'utilisation de liens non orientés. Dans le réseau valué, le lien entre un agent i et un agent j est d'intensité k si l'agent i et l'agent j participe à exactement k projets ensemble.

Nous construisons la matrice de relations entre ces agents X^V . La matrice X^V est une matrice carrée de taille $n \times n$. Un élément de X^V est noté x_{ij}^V avec $i, j \in N$. L'élément x_{ij}^V de la matrice X^V a la valeur k si l'agent i et l'agent j participent à k projets ensemble.

Dans le cas où $i \neq j$:

$$x_{ij}^V = \begin{cases} k & \text{si } i \text{ et } j \text{ participent à } k \text{ projets ensemble} \\ 0 & \text{si l'agent } i \text{ et l'agent } j \text{ n'ont aucun projet en commun} \end{cases}$$

et si $i = j$ (les éléments de la diagonale) :

$$x_{ij}^V = x_{ii}^V = \text{le nombre de projets auxquels participe l'agent } i$$

La matrice X^V s'obtient à partir de la matrice d'affiliation A utilisée pour les “two-mode networks” :

$$X^V = A.A'$$

avec A la matrice d'affiliation et A' la transposé de la matrice d'affiliation. Lorsque l'on construit le réseau, on recode l'ensemble des éléments de la diagonale à 0. Le nombre de liens du réseau est égal à :

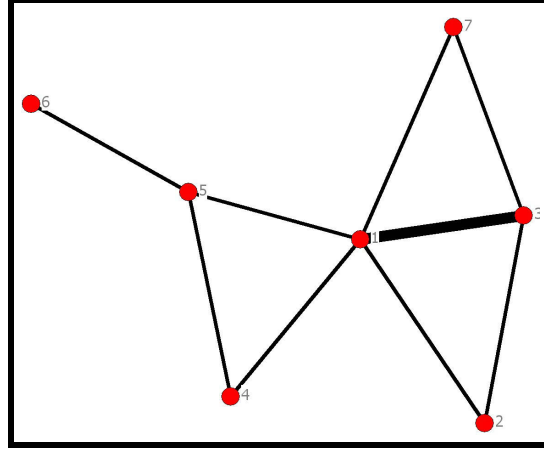
$$l^V = |L^V| = \sum_{i < j, i \neq j} x_{ij}^V$$

Si l'on reprend l'exemple 1 de la section 1.2.1, on obtient :

Agents	Ag. 1	Ag. 2	Ag. 3	Ag. 4	Ag. 5	Ag. 6	Ag. 7
Agent 1	3	1	2	1	1	0	1
Agent 2	1	1	1	0	0	0	0
Agent 3	2	1	2	0	0	0	1
Agent 4	1	0	0	1	1	0	0
Agent 5	1	0	0	1	2	1	0
Agent 6	0	0	0	0	1	1	0
Agent 7	1	0	1	0	0	0	1

$Ag.^* = Agent$

Le graphe associé à cette matrice valuée est représenté de la façon suivante :



Graphe 4 : Représentation du réseau des agents en “one-mode network” à partir de l'exemple 1.

Lorsque le lien entre i et j est différent de 0 alors il existe un lien direct entre i et j , c'est-à-dire que i et j communiquent directement sans avoir besoin d'intermédiaire.

L'intérêt de cette matrice est qu'elle rend compte à la fois de l'existence d'un lien entre l'agent i et l'agent j et de l'intensité de ce lien. Les réseaux valués permettent d'étudier la redondance des collaborations dans le réseau. Dans le cadre des PCRDT, cette intensité définit le nombre de fois où deux agents se retrouvent dans un même projet. Dans la section 2.3 du chapitre 2, nous montrerons que cette représentation est la seule utilisable pour étudier les réseaux de régions au sein des PCRDT.

Il est possible de synthétiser les informations ci-dessus pour obtenir une matrice qui rend compte uniquement de l'existence d'un lien entre deux agents. De cette manière, nous obtenons une matrice binaire X^B . Cette matrice s'obtient à partir de X^V de la manière suivante :

$$x_{ij}^B = \begin{cases} 1 & \text{si } i \neq j \text{ et si } x_{ij}^V \geq 1 \\ 0 & \text{sinon} \end{cases}$$

Le nombre de liens du réseau est égal à :

$$l^B = |L^B| = \sum_{i < j} x_{ij}^B$$

A partir de l'exemple 1, on obtient la matrice suivante :

Agents	Ag. 1	Ag. 2	Ag. 3	Ag. 4	Ag. 5	Ag. 6	Ag. 7
Agent 1	0	1	1	1	1	0	1
Agent 2	1	0	1	0	0	0	0
Agent 3	1	1	0	0	0	0	1
Agent 4	1	0	0	0	1	0	0
Agent 5	1	0	0	1	0	1	0
Agent 6	0	0	0	0	1	0	0
Agent 7	1	0	1	0	0	0	0

$$Ag.^* = Agent$$

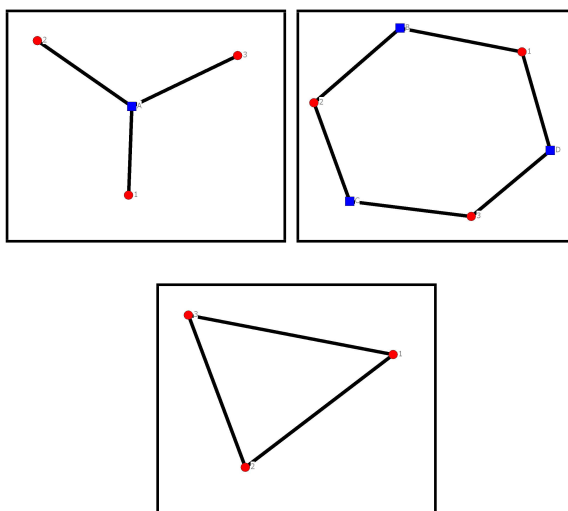
Pour étudier les réseaux de collaborations dans les PCRDT et vérifier si les PCRDT ont atteint leurs objectifs, il est intéressant d'utiliser les deux représentations. Chacune amène des informations différentes et complémentaires. Par exemple, lorsque l'on étudie la position d'un agent au sein du réseau, il sera ainsi possible d'évaluer cette position à partir du nombre de collaborations totales de l'agent mais aussi à partir du nombre de partenaires distincts de chaque agent. De même, nous montrerons que l'étude du réseau dans sa globalité apporte des éléments différents selon que l'on étudie le graphe binaire ou le graphe valué.

La représentation du réseau en "one-mode network" plutôt qu'en "two-mode network" simplifie l'analyse puisqu'on s'intéresse seulement à un ensemble, soit celui des agents, soit celui des projets. De cette manière, cela rend le graphe plus lisible et sa compréhension plus facile. Ainsi, entre les agents participants aux PCRDT, les projets n'apparaissent pas comme intermédiaires, rendant la lecture des liens entre agents plus directs.

Cette représentation permet également de réaliser des comparaisons entre deux périodes différentes et/ou deux secteurs différents. Ainsi, dans le cadre des PCRDT, les agents peuvent participer à plusieurs programmes cadres dans différents domaines et à différents moments. Il est donc possible de retrouver d'un secteur à un autre et/ou d'une période à une autre un même agent contrairement à un projet qui correspond à un PCRDT et à une thématique.

Toutefois, il ne faut pas éluder le fait que la représentation en "one-mode network" conduit à une perte d'information contrairement aux "two-mode network". En effet, lorsque l'on représente uniquement les réseaux de collaborations avec les agents, on ne connaît pas quels sont les projets qui lient

les agents entre eux. Par exemple, si trois agents sont directement connectés dans un “one-mode network”, il est impossible de déterminer si ces agents appartiennent tous les trois au même projet ou s'ils sont liés grâce à des projets différents. Prenons l'exemple suivant :



Graphes 5, 6 et 7 : Illustrations de la perte d'information lors du passage des “two-mode networks” aux “one-mode networks”.

Sur cet exemple, on s'aperçoit clairement de la perte d'information qu'engendre la représentation en “one-mode network”. Les deux graphes du haut sont des “two-mode networks” qui représente chacun une situation bien distincte. Pourtant, lorsque l'on souhaite représenter le réseau de collaborations entre agents de ces deux réseaux, on obtient une représentation strictement identique pour les deux réseaux. La représentation sous forme de graphe valué permet en partie de limiter cette perte d'information en indiquant le nombre de projets en commun que possèdent deux agents mais pas le nom de ces projets.

Lorsque le nombre d'agents au sein des projets est faible, cette hypothèse est raisonnable. C'est pour cela notamment, que lorsqu'ils étudient les réseaux de co-auteurs, les auteurs restreignent leur étude à des disciplines où le nombre de co-auteurs par article est relativement faible comme en Économie et, même parfois, ils ne conservent que les articles dont le nombre de co-auteurs est faible que ce soit pour des raisons techniques (Goyal, van der Leij et Moraga-Gonzalez, 2006) ou pour éviter de construire des cliques trop importantes (Fafchamps, van der Leij et Goyal, 2010). Lorsque le nombre d'agents au sein des projets est élevé, comme c'est le cas des projets des

PCRDT, l'hypothèse suivant laquelle tous les agents au sein d'un même projet collaborent deux à deux est plus discutable. Cette limite doit être prise en considération dans l'interprétation des résultats obtenus en utilisant cette représentation.

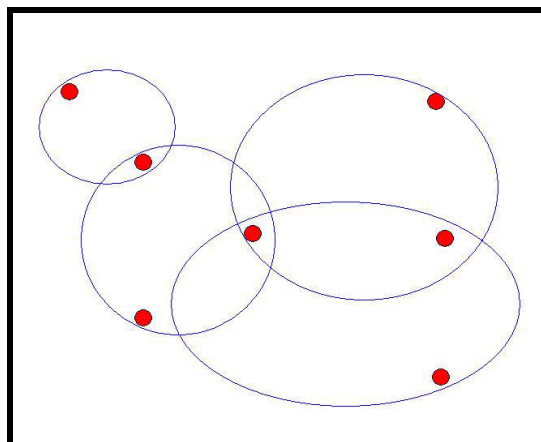
En résumé, le choix entre les deux représentations, entre “one-mode network” et “two-mode network”, dépend en premier lieu de l'objectif que l'on cherche à atteindre. Ces deux représentations apportent des informations différentes et sont donc complémentaires. L'analyse en “two-mode network” peut notamment être vue comme une première étape avant d'utiliser la représentation en “one-mode network”. Il ne faut pas oublier que ce choix n'est pas neutre et qu'il peut entraîner des différences en termes d'analyse. Billand, Frachisse et Massard (2008) montrent à partir des indicateurs réseaux que l'analyse que l'on peut faire des PCRDT et les conclusions auxquelles on arrive peuvent être différentes selon le choix de la représentation, plus particulièrement lorsque l'analyse se fait à un niveau individuel (recherche des agents les plus centraux du réseau par exemple).

1.2.3 Hypergraphes : des représentations fidèles mais incompatibles avec l'analyse réseau

Une autre manière de représenter les réseaux de collaborations des PCRDT est d'utiliser des hypergraphes. Les hypergraphes peuvent apparaître comme la représentation la plus fidèle des réseaux de collaborations. En effet, tous les agents qui participent à un projet sont représentés dans un même sous-ensemble sans que la représentation nous oblige à faire des hypothèses sur les relations entre les agents dans ce sous-ensemble. Les hypergraphes autorisent les agents à appartenir à plusieurs sous-ensembles simultanément.

De manière formelle, un hypergraphe consiste en un ensemble d'objets (dans le cas des PCRDT, ce sont les agents) et dans une collection de sous-ensemble d'objets (les projets), tels que chaque objet appartient à au moins un sous-ensemble et tel qu'aucun sous-ensemble n'est vide (Berge, 1989). Un hypergraphe est noté $H(N, L)$.

Si l'on cherche à construire l'hypergraphe associé à l'exemple 1, nous obtenons la représentation suivante :



Graphe 8 : Représentation du réseau de l'exemple 1 sous forme d'hypergraphe.

Cette représentation est d'une certaine façon celle qui correspond le mieux au réseau d'affiliation car on ne suppose aucun lien entre les agents au sein d'un sous-ensemble. Cette représentation est visuellement plus pertinente puisque l'aspect ensembliste des projets apparaît clairement. Toutefois, cette représentation convient peu aux bases de données qui comportent un nombre important de collaborations. La représentation des réseaux devient rapidement ardue et l'analyse complexifiée. De plus, cette représentation ne permet pas d'utiliser les indicateurs couramment usités dans l'analyse des réseaux sociaux (cf. sections 1.3, 1.4 et 1.5). De fait, nous ne retenons pas cette représentation pour la suite.

Après avoir déterminé quelle était la représentation la plus judicieuse pour étudier les PCRDT, nous présentons dans les deux sections suivantes un certain nombre d'indicateurs pour mener à bien l'analyse. L'analyse réseau porte sur deux niveaux différents : le premier, d'ordre macroéconomique, consiste à étudier les propriétés générales du réseau et le second, d'ordre microéconomique, consiste à étudier les propriétés individuelles des sommets qui composent le graphe. Nous présentons tout d'abord les notions et les indicateurs nécessaires à l'étude des propriétés générales du réseau.

1.3 Propriétés générales des réseaux

1.3.1 Densité d'un réseau

L'observation du nombre de liens et de la densité du réseau permettent d'étudier la connectivité entre les agents. La densité d'un graphe est égale au

rapport entre le nombre de liens réalisés et le nombre de liens potentiellement réalisables. Plus le nombre de liens est important (à nombre d'agents donnés), plus la connectivité du réseau est forte. La densité du réseau se calcule différemment selon la représentation utilisée. Soient n le nombre d'agents et l le nombre de liens du réseau étudié. Le tableau ci-dessous donne la densité du réseau (du graphe) dans le cas du réseau valué et dans le cas du réseau binaire.

Indicateurs	Réseau valué	Réseau binaire
Nombre d'agents	n	n
Nombre de liens (ou valeur totale des liens)	l^V	l^B
Nombre de liens potentiels	$\frac{n \times (n-1)}{2}$	$\frac{n \times (n-1)}{2}$
Densité (Δ)	$\Delta^V = \frac{l^V}{\frac{n \times (n-1)}{2}} \times 100$	$\Delta^B = \frac{l^B}{\frac{n \times (n-1)}{2}} \times 100$

Tableau 1 : Calcul de la densité des différents réseaux.

La densité d'un graphe permet de savoir si les agents du graphe sont fortement connectés entre eux ou non. Plus la valeur obtenue est proche de 100, plus on se rapproche du graphe complet (le nombre maximal de liens correspond au nombre de liens du graphe complet). Cette mesure est utilisée pour étudier la cohésion au sein d'un réseau et peut aussi l'être pour étudier la cohésion d'un sous-groupe. Toutefois, le calcul et l'interprétation sont différents selon le réseau étudié.

Dans le cas des réseaux binaires ("one-mode network"), la densité correspond au pourcentage de liens effectivement créés dans le réseau. Le nombre maximal de liens correspond à la situation où tous les agents collaborent ensemble. Ainsi, on peut caractériser, pour chaque domaine technologique, la plus ou moins grande densité du réseau même si on ne sait pas si c'est au travers de nombreux projets de faible taille et/ou au travers de grands projets contenant un nombre important d'agents.

Dans le cas d'un réseau valué ("one-mode network"), la mesure de densité est plus complexe dans le sens où à un même indicateur peuvent correspondre trois indicateurs différents :

- La première Δ_l^V est la densité présentée dans le tableau. Elle est égale au rapport entre la somme des valeurs de tous les liens du réseau (l^V) et le nombre de liens maximales. Cette mesure permet de comparer deux graphes de tailles différentes.
- La seconde Δ_2^V rapporte l^V au nombre de liens du graphe binaire l^B . Cette mesure donne l'intensité moyenne des liens qui composent le graphe.

$$\Delta_2^V = \frac{l^V}{l^B}$$

- La troisième Δ_3^V est plus complexe. Dans un premier temps, il faut déterminer le lien qui est le plus intense. Cette valeur est alors considérée par défaut comme la valeur maximale d'un lien entre deux agents du graphe. On cherche :

$$v_{max} = \text{Max}\{x_{ij}^V\} \text{ pour } i, j \in N \text{ et } i \neq j$$

Au maximum un agent ne peut donc avoir que $n - 1$ liens d'une valeur maximale chacun de v_{max} . La valeur totale des liens du graphe ne peut donc excéder

$$\frac{n \times (n - 1) \times v_{max}}{2}$$

On divise alors par deux car le réseau est symétrique et $x_{ij}^V = x_{ji}^V$. Au final, pour obtenir la densité, on rapporte l^V au nombre obtenu ci-dessus :

$$\frac{l^V}{\frac{n \times (n - 1) \times v_{max}}{2}}$$

La densité ainsi obtenue donne le pourcentage des liens qu'ont effectivement réalisé les agents en prenant en compte le fait qu'il leur est possible de réaliser plusieurs fois le même lien avec le même agent.

Afin d'illustrer ces mesures, nous les appliquons à l'exemple 1.

Indicateurs	Réseau valué	Réseau binaire
Nombre d'agents	7	7
Nombre de liens (ou valeur total des liens)	10	9
Nombre maximal de liens	21	21
Densité	$\Delta_1 = 47,6\% = \frac{10}{21}$ $\Delta_2 = 1,1 = \frac{10}{9}$ $\Delta_3 = 23,8\% = \frac{10}{42}$	$42,9\% = \frac{9}{21}$

Tableau 2 : Calcul de la densité pour l'exemple 1.

Dans le cas d'un réseau valué, le choix entre ces différentes mesures se fait en fonction de ce que l'on souhaite montrer. La dernière mesure demande de faire une hypothèse sur le nombre maximal de fois qu'un agent peut collaborer avec un autre agent ce qui n'est pas forcément conforme à la réalité. C'est la raison pour laquelle, il est préférable d'utiliser soit la première soit la seconde mesure. Une autre possibilité consiste à mesurer la densité du graphe en ne considérant que les liens de valeur supérieure à un certain seuil. Par exemple, on peut calculer la densité du réseau de collaborations en ne prenant en compte que les liens de force 2 ou plus. Ce seuil est particulièrement intéressant puisqu'il permet de déterminer le pourcentage de liens qui correspondent à des collaborations redondantes. Ces collaborations peuvent être interprétées comme une relation de confiance ou un partenariat privilégié entre deux agents. Plus le seuil que l'on fixe est élevé plus on élimine des liens. De cette manière, on peut étudier si en plus d'être dense (nombreux liens distincts entre les agents) le réseau et aussi intensément connecté (nombreux liens entre deux agents).

1.3.2 Composantes et singletons

Les deux notions que l'on présente dans ce paragraphe renvoient à la connectivité du graphe. Nous cherchons notamment à déterminer si le graphe est connexe, c'est-à-dire si tous les agents du réseau peuvent communiquer entre eux à l'aide de liens directs et/ou indirects. Si au moins un agent n'est pas relié de façon directe ou indirecte à un autre agent alors le graphe est déconnecté. Dans le cas des PCRDT, il est important de s'intéresser à la connectivité du graphe. De plus l'Union Européenne a pour objectif de construire un EER qui se veut un espace d'échange et de collaborations intégré au sein de l'UE. En effet, dans le cadre de collaborations en R&D, il est important que les agents soient connectés les uns aux autres afin que l'information, les connaissances et les externalités circulent bien au sein du

réseau et améliore ainsi l'efficacité de la recherche et donc l'innovation. En d'autres termes, cet EER sera d'autant plus efficace que le réseau formé par les PCRDT sera connexe.

Une composante est un sous-ensemble du graphe tel que tous les agents d'une même composante sont reliés entre eux (directement ou indirectement) et aucun agent de la composante n'est lié à un autre élément qui n'est pas dans la composante. Une composante est un sous-graphe particulier du graphe. Si le graphe est connexe alors, il possède une seule composante. En observant le nombre de composantes et l'évolution de ce nombre dans le temps, on observe si l'espace de recherche tend ou non à devenir un espace intégré et connexe où chaque agent est en relation avec les autres agents au moins de manière indirecte. Cet indicateur est souvent complété par une mesure de la taille de la plus grande composante, la taille étant mesurée par le nombre d'agents ou de liens présents au sein de la grande composante. Il est aussi possible de réaliser des comparaisons sectorielles. Dans un réseau de collaborations, il peut être important que le réseau possède le moins de composantes possibles afin d'améliorer la diffusion de l'information, des connaissances et des externalités.

Un singleton est un agent du graphe qui ne possède aucun lien au sein du graphe. Dans un réseau de collaborations, il faut éviter d'avoir des singletons car ce sont autant d'agents qui ne sont pas intégrés au réseau et qui ne profitent pas des connaissances créées au sein de ce réseau. De la même manière, les connaissances développées par un agent singleton ne se diffusent pas au reste du réseau. Dans le cas des PCRDT, les projets sont de type collaboratifs. Les projets appartenant à une des 7 priorités thématiques sont composés (sauf exception) d'au moins deux agents (de nombreux projets doivent même comporter au minimum trois agents), il est donc peu probable de voir apparaître des singletons.

1.3.3 Distance et diamètre

Les indicateurs relatifs à la distance et au diamètre d'un graphe permettent d'observer comment les agents sont connectés les uns aux autres. Ces indicateurs permettent alors d'estimer dans quelle mesure les informations circulent rapidement entre les agents d'un même réseau de collaboration en R&D. Ces informations peuvent notamment concerner les caractéristiques des partenaires et être stratégiques pour des collaborations futures. Ainsi, on peut supposer que même des agents qui ne sont pas directement connectés peuvent communiquer ou accéder à de l'information l'un sur l'autre via leurs partenaires communs. Ce type d'information peut jouer un rôle important dans les choix de collaborations futures. Différents types d'informations

transitent dans le réseau formé par les PCRDT :

- Connaissances initiales des agents ;
- Connaissances créées dans le cadre des projets se déroulant au sein des PCRDT ;
- Connaissances sur les partenariats ;
- Connaissances sur les marchés, technologies, produits, . . . ;
- Etc.

Ces informations circulent dans le réseau, d'autant plus qu'au sein des projets de R&D, les agents ne maîtrisent pas entièrement la diffusion du savoir. Une partie se diffuse dans le réseau à leur insu. On parle d'externalités de connaissance. Un agent peut ainsi bénéficier de la recherche et de la production de connaissances générées par un projet auquel il ne participe pas.

Il existe plusieurs hypothèses sur la manière dont circulent les informations au sein d'un réseau. La première consiste à supposer que la circulation des connaissances et des informations est parfaite. Dans ce cas, la distance compte peu. La seconde considère que, dans un réseau, un agent profite des externalités de manière inversement proportionnelle à la distance qui le sépare de la source (hypothèse de "decay", (Bala et Goyal, 2000), c'est la raison pour laquelle on peut supposer que plus la distance moyenne est faible, plus la circulation des informations est satisfaisante. Pour un agent, il est d'autant plus facile de capter ces externalités de connaissance que celui-ci est situé à proximité de la source. Cette remarque s'applique aussi pour les externalités. La dernière hypothèse que l'on peut faire est qu'il existe une borne (une distance) au sein du réseau telle que si un agent est situé à une distance supérieure à cette borne, alors il ne profite pas des connaissances créées ("bounded network") (Jackson et Wolinsky, 1996). Les connaissances créées au sein d'un projet sont censées circuler parfaitement au sein de celui-ci. Dans le cas des réseaux de collaborations, la circulation de l'information n'est pas parfaite. Les connaissances issues des collaborations au sein d'un projet ne sont qu'imparfaitement transmises aux agents qui ne participent pas au projet dans le cadre de la participation à d'autres projets. Dans ces réseaux, des entreprises et des centres de recherche collaborent mais il ne faut pas oublier que par la suite ils ont des objectifs de profit ou scientifiques qui font que ces agents n'ont pas intérêt à diffuser toutes leurs connaissances. Les hypothèses les plus crédibles sont les hypothèses dites de "decay" et "bounded network", hypothèses qui peuvent se cumuler, d'où l'intérêt pour un agent d'être situé à faible distance des autres agents. Pour faciliter la diffusion des connaissances et des informations, l'UE a intérêt à ce que le réseau des PCRDT soit caractérisé par une distance moyenne et un diamètre faibles, synonyme d'un espace intégré où la recherche des uns se diffuse et bénéficie aux autres agents.

Pour présenter les notions de distance et de diamètre de manière formelle, nous nous plaçons dans le cadre le plus simple à savoir un réseau où l'on retient seulement la formation des liens (cas binaire). La distance géodésique entre l'agent i et l'agent j est notée d_{ij} . Dans la mesure où nous sommes dans le cas d'un réseau non orienté et symétrique nous avons $d_{ij} = d_{ji}$. La distance maximale possible entre deux agents dans un graphe connexe est de $n - 1$. Elle correspond au cas où le graphe aurait la forme d'une ligne.

Lorsque deux agents n'appartiennent pas à une même composante, c'est-à-dire qu'ils ne possèdent ni lien direct ni lien indirect entre eux, on considère qu'ils ne peuvent pas communiquer et qu'on ne peut pas définir la distance entre ces deux agents. Par la suite, pour faciliter le calcul, on fera l'hypothèse que la distance entre deux agents non liés directement ou indirectement est égale à infini $d_{ij} = +\infty$. Certains logiciels (Ucinet par exemple) considèrent que la distance entre deux agents qui n'appartiennent pas à la même composante est de $(n - 1) + 1 = n$ soit la distance maximale possible plus un. Cette solution ne nous semble pas optimale dans la mesure où on attribue une distance entre deux agents certes fortes mais qui ne reflète pas la réalité. Dans le cas où le graphe n'est pas connexe, on calcule en général la distance moyenne au sein la plus grande composante si celle-ci est jugée de taille suffisante.

Dans le cas d'un graphe binaire composé uniquement d'agents, la matrice qui résume l'ensemble de ces distances est notée D^B . Les éléments de la diagonale ne sont pas renseignés. La distance moyenne d'un réseau, notée \bar{d} , est calculée de la manière suivante :

$$\bar{d} = \sum_{i < j} d_{ij}$$

Enfin, le diamètre d'un graphe, noté D , est le chemin géodésique le plus long du graphe. Il correspond à :

$$D = \text{Max}_{i,j} \{d_{ij} \text{ pour } i, j \in N \text{ et } i \neq j\}$$

Ces deux indicateurs peuvent être calculés si et seulement si le graphe est connexe. Dans le cas contraire, il faudra réaliser une approximation en calculant ces indicateurs sur la plus grande composante. L'approximation sera d'autant plus juste que la plus grande composante comprend un nombre important d'agents. Pour présenter ces indicateurs, nous nous sommes placés dans le cadre le plus simple. Ces indicateurs se calculent de manière identique lorsque l'on considère un réseau valué.

Remarque : Dans le cas d'un réseau valué, il est possible de calculer la distance moyenne et le diamètre d'un réseau pour différents seuils d'intensité. Si l'on compare la distance moyenne et le diamètre du réseau dans sa globalité et ceux du réseau où l'on ne considère que les liens d'intensité supérieure ou égale à 2 et que l'on trouve un écart important, alors on peut considérer que les liens de faible intensité sont essentiels pour permettre la diffusion des connaissances. Cela peut s'apparenter à une équivalence du concept de Granovetter pour les réseaux valués sur la force des liens faibles dans le cas des réseaux binaires (Granovetter, 1973, 1983). Dans le cas contraire, alors cela signifie que les liens de forte intensité suffiraient à assurer une diffusion efficace des connaissances et des informations au sein du réseau.

1.3.4 Coefficient de cluster

Le coefficient de cluster permet de mesurer si les partenaires d'un agent possèdent aussi des liens entre eux. Au niveau du graphe (au niveau global), cet indicateur permet de déterminer quel est le degré de connexion entre les agents. Plus le graphe a un coefficient de cluster élevé, plus il présente une forte connexion entre les agents. Au niveau individuel (niveau local), il permet de repérer la formation de sous-groupes d'agents fortement connectés les uns aux autres. De nombreuses études ont montré que dans les réseaux réels, plus particulièrement dans les réseaux sociaux, les agents tendent à créer des groupes de faible taille fortement connectés (Holland and Leinhardt, 1971 ; Watts et Strogatz, 1998). Pour mettre en évidence ce phénomène, les études comparent les réseaux observés avec des réseaux de type aléatoire.

Pour obtenir le coefficient de cluster du réseau, il faut dans une première étape déterminer le coefficient de cluster de chaque agent. Le coefficient de cluster d'un agent est noté CCL_i . Dans le cas d'un réseau binaire, il s'obtient de la manière suivante :

$$CCL_i = \frac{\sum_{j,k \in N_i} x_{jk}^B}{n_i(n_i - 1)}$$

avec N_i l'ensemble des partenaires de l'agent i ($N_i \subset N$), n_i le nombre de partenaires de i et x_{jk}^B l'élément de la matrice d'adjacence X associé à la relation entre l'agent j et l'agent k . A partir de cet indicateur, le coefficient de cluster du graphe s'obtient de la manière suivante :

$$CCL = \frac{1}{n} \sum_{i \in N} CCL_i$$

Dans le cas des réseaux d'affiliation ou "bipartite graph", utiliser cette formule introduirait un biais. En effet, par construction et étant donné l'hypothèse qui nous permet de passer des "two-mode networks" aux "one-mode networks", tous les projets forment un sous-graphe complet. De fait, un agent qui participe à un seul projet a obligatoirement un coefficient de cluster égal à 1. Par conséquent, utiliser la formule ci-dessus conduit à surestimer le coefficient de cluster et donc la connectivité du graphe. Pour pallier à cette lacune, Newman, Watts et Strogatz (2001) proposent la formule suivante :

$$CCl_{bis} = \frac{3 \cdot \text{number of triangles on the graph}}{\text{number of connected triples of vertices}}$$

Le nombre de triangles dans le réseau est ici défini comme le nombre de trios d'agents où chaque agent est connecté aux deux autres et le terme "triples" renvoie à la situation où au moins un des trois agents est connecté aux deux autres. Le nombre 3 en facteurs s'explique par le fait que chaque triangle compte pour 3 "triples", un pour chaque agent du triangle. De cette manière, le CCl_{bis} est compris entre 0 et 1. Grâce à cet indicateur, il est ainsi possible de contourner le biais introduit par notre hypothèse de départ.

Le coefficient de cluster permet de compléter les propriétés générales du réseau. Une fois que l'on possède toutes les propriétés du réseau, il est possible d'établir si le réseau possède une structure particulière. Il est intéressant de déterminer notamment si les réseaux de collaborations des PCRDT ne se rapprochent pas des réseaux en "petit-monde" (small world networks). Ces réseaux sont présentés comme ayant une architecture facilitant la transmission d'informations au sein du réseau par l'intermédiaire d'un nombre minimal de liens et sont aussi caractérisés par un coefficient de cluster élevé. Nous reviendrons plus en détail sur ce type de réseau au moment d'aborder l'étude des propriétés générales des réseaux PCRDT pour les deux domaines technologiques choisis.

Nous venons de présenter plusieurs indicateurs qui permettent d'étudier les propriétés générales du réseau. Dans la sous-section suivante, nous nous intéressons aux indicateurs qui permettent d'évaluer le rôle d'un agent au sein du réseau.

1.4 Analyse des réseaux au niveau individuel

Dans cette section, nous présentons un certain nombre d'indicateurs qui mesurent le rôle joué par les agents au sein du réseau. Ces indicateurs per-

mettent de mettre en avant les agents qui possèdent des propriétés particulières favorables à la connectivité du réseau (nombre élevé de liens par exemple) et à son efficacité (liaisons entre les agents). Les PCRDT cherchent à attirer dans chaque thématique les acteurs principaux de la recherche en Europe (leaders industriels dans les domaines technologiques financés par exemple). De plus, dans l'objectif de créer un EER et de favoriser la création de nouvelles connaissances, il est intéressant d'avoir des agents qui participent à plusieurs projets dans des domaines technologiques différents afin de favoriser ou d'établir des croisements surtout dans un contexte où la recherche est de plus en plus pluri-disciplinaire. Les indicateurs permettent de mettre en avant les agents les plus participatifs ainsi que ceux qui servent de passerelle entre les domaines technologiques. Notons que les indicateurs retenus ne prennent pas en compte les caractéristiques des agents (chiffres d'affaires, type d'agents,...), mais prennent en compte, dans le calcul de la centralité, seulement leur position au sein du réseau. Il est ainsi possible d'identifier les agents importants du réseau sans a priori.

En fait, il existe un nombre important de définitions de la centralité dans la littérature sur les réseaux sociaux (Faust, 1997 ; Borgatti et Everett, 1997). Dans cette section, nous présentons les indicateurs qui nous semblent les plus pertinents pour étudier la centralité des agents qui participent aux PCRDT.

1.4.1 La pertinence des indicateurs de centralité les plus usités

De nombreuses études portant sur les réseaux de collaborations dans les PCRDT (Breschi et Cusmano, 2004 ; Malerba et al., 2006 ; Røediger-Schluga et Barber, 2007) ou sur d'autres formes de collaborations comme les réseaux de co-auteurs par exemple (Cassiman et Veugelers, 2002, 2002 ; Goyal, van der Leij et Moraga-Gonzalez, 2006) cherchent à déterminer, à l'aide de différents indicateurs, quels sont les agents les plus centraux. Chaque indicateur correspond à une définition de la centralité bien précise (Faust, 1997). Dans le cadre de notre travail, nous retenons les définitions suivantes de la centralité :

- Un agent est central s'il est actif à l'intérieur du réseau. Afin d'observer l'activité d'un agent, nous utilisons la mesure de degré (Freeman, 1979).
- Un agent est central s'il peut communiquer ou interagir rapidement avec l'ensemble des agents du réseau. Afin de mesurer cette forme de centralité, nous calculons la centralité de proximité ou "closeness centrality" (Freeman, 1979).
- Un agent est central s'il est indispensable au bon fonctionnement du réseau. Plus précisément, un agent est central s'il est un intermédiaire

pour les autres agents du réseau. L'indicateur associé à cette mesure est la centralité d'intermédiarité ou "betweenness centrality" (Freeman, 1979).

- Un agent est central s'il est lié avec des agents qui sont eux-mêmes centraux. Nous calculons la centralité pour les valeurs propres ou "eigenvector centrality" (Bonacich, 1972a).

Remarque : Dans la thèse, ces indicateurs sont définis dans le cas des "one-mode networks". L'article de Faust (1997) montre la manière de les adapter au cas des "two-mode networks". Pour chaque indicateur, nous proposerons une normalisation de cet indicateur qui est une mesure qui permet de comparer la position d'un agent dans deux réseaux de taille différente (en termes de nombre d'agents), dans deux PCRDТ ou dans deux domaines technologiques différents par exemple. La normalisation permet aussi de comparer le rôle joué par deux agents différents dans des réseaux distincts. Dans le cas des "two-mode networks", Borgatti et Everett (1997) proposent une adaptation de cette normalisation. Même si nous n'utilisons que la représentation en "one-mode network", il est important de préciser que ces indicateurs s'appliquent aux différentes représentations.

Dans les paragraphes suivants, nous présentons les différents indicateurs de manière plus détaillée.

1.4.1.1. Centralité de degré ou "degree centrality"

La centralité de degré (ou degré) est la mesure de centralité la plus simple. Le degré d'un agent dans un réseau est égal à son nombre de liens. Un agent est considéré d'autant plus actif qu'il possède des liens dans le réseau. Plus l'agent a de liens, plus il est considéré comme central. Cet indicateur peut à la fois rendre compte du niveau d'activité d'un agent et peut être interprété en terme de diffusion de connaissances, à savoir quelle est la capacité d'un agent à diffuser des connaissances mais aussi à bénéficier des connaissances des autres agents.

Dans un réseau de collaborations comme celui des PCRDТ, une manière d'interpréter le degré en termes de diffusion de connaissances consiste à considérer que les connaissances ou les processus de diffusion de savoirs impliquent que seuls les liens directs permettent de transférer le savoir. Les agents qui ne sont pas impliqués dans la relation ne peuvent bénéficier des connaissances. Enfin, le degré mesure la capacité d'un agent à s'intégrer au réseau de recherche, ainsi qu'à bénéficier des connaissances présentes dans le réseau en créant des liens directs grâce à des partenariats.

La signification du degré de l'agent change selon que l'on est dans le cadre valué ou binaire. Dans le cas d'un réseau valué, le degré d'un agent correspond au nombre de liens que possède l'agent dans le graphe en prenant en compte les redondances des liens tandis que dans le cas d'un réseau binaire, le degré donne le nombre de partenaires distincts que possède un agent. Il est important de noter la complémentarité des informations révélées par la centralité de degré dans ces deux représentations des PCRDT.

Le degré d'un agent dans un graphe valué est noté $CD^V(i)$ pour $i \in N$. Par la suite, les indicateurs qui seront calculés dans le cas du réseau valué seront notés avec un exposant V et ceux calculés dans le cadre du réseau binaire seront notés avec l'exposant B . Cet indicateur est obtenu de la manière suivante :

$$CD^V(i) = \sum_{j \neq i} x_{ij}^V$$

pour $i \in N$ et $j \in N$

Pour chaque agent, il est possible d'avoir un indicateur normalisé qui prend en compte la taille du réseau dans lequel l'indicateur est mesuré. Cela permet de comparer la position d'un agent dans deux réseaux différents. L'indicateur normalisé s'obtient de la façon suivante.

$$CD^V(i)' = \frac{\sum_{j \neq i} x_{ij}^V}{n - 1}$$

pour $i \in N$ et $j \in N$, $n - 1$ étant le nombre maximum d'agents avec qui l'agent i peut avoir des liens. Le vecteur qui résume le degré obtenu pour chaque agent du réseau est noté CD_N^V . On a :

$$CD^V = X^V \cdot V_1$$

avec X^V la matrice de relations entre les agents et V_1 un vecteur colonne de taille $(n \times 1)$ dont tous les éléments sont égaux à 1.

De la même manière, on peut définir cet indicateur dans le cadre d'un réseau binaire en utilisant la matrice d'adjacence X^B au lieu de X^V .

L'indicateur ainsi obtenu est noté $CD^B(i)$. Pour illustrer l'ensemble de ces indicateurs, nous reprenons l'exemple 1.

Indicateur	$CD^V(i)$	$CD^V(i)'$	$CD^B(i)$	$CD^B(i)'$
Agent 1	6	$\frac{6}{6}$	5	$\frac{5}{6}$
Agent 2	2	$\frac{2}{6}$	2	$\frac{2}{6}$
Agent 3	4	$\frac{4}{6}$	3	$\frac{3}{6}$
Agent 4	2	$\frac{2}{6}$	2	$\frac{2}{6}$
Agent 5	3	$\frac{3}{6}$	3	$\frac{3}{6}$
Agent 6	1	$\frac{1}{6}$	1	$\frac{1}{6}$
Agent 7	2	$\frac{2}{6}$	2	$\frac{2}{6}$

Tableau 3 : Indicateur de degré pour les agents de l'exemple 1

Le degré est une mesure facile à calculer et intuitive. Le degré, outre permettre de comparer la place des agents au sein du réseau, est directement interprétable. Le problème est que dans le cas des PCRDT, cette mesure est influencée par l'hypothèse que l'on fait sur les collaborations au sein de chaque projet. Ainsi, si un agent participe à un seul projet mais un projet de grande envergure avec de nombreux agents, il aura un degré élevé qui ne reflète pas pour autant une participation importante aux PCRDT. Le fait d'utiliser des mesures valuées diminue un peu ce biais. C'est pourquoi en plus du degré, nous calculerons le nombre de projets auquel participe chaque agent.

Le degré cependant est une mesure faiblement discriminante. Deux agents peuvent avoir un degré identique sans pour autant être dans des positions symétriques au sein du réseau. Par exemple, si un agent possède deux liens au sein d'une composante de 20 agents et un autre possède deux liens dans une composante de 3 agents, ces deux agents sont aussi centraux l'un que l'autre selon le degré alors que l'agent du premier cas semble avoir accès à plus d'agents et donc à plus de connaissances et d'informations que celui du second. Pour que les deux agents soient aussi centraux, il faudrait supposer que, dans le réseau, les informations et les externalités ne circulent que dans les liens directs et ne peuvent s'étendre au reste du réseau. On se situerait alors dans un environnement peu propice à l'innovation. Dans l'exemple donné ci-dessus, il est difficile de dire que les deux agents occupent une po-

sition strictement équivalente¹. Le principal défaut du degré est qu'il est calculé de manière locale et tient seulement compte du voisinage de l'agent. Le degré reste toutefois une première mesure intéressante à calculer mais pas suffisante pour caractériser la place d'un agent dans le réseau.

Pour compléter la présentation de l'indicateur de degré, il est possible de calculer un indicateur de degré pour le réseau et non pour l'un des agents. Cet indicateur permet de comparer le réseau étudié à un réseau de référence qui est le réseau en forme d'étoile. Grâce à l'indicateur de centralité de degré du réseau, nous obtenons une information supplémentaire sur la forme du réseau en plus des informations obtenues grâce aux indicateurs généraux vus à la section 1.3. En comparant le réseau étudié avec un réseau en forme d'étoile de même taille, on peut ainsi déduire si le réseau présente une distribution des liens hétérogènes (avec des agents qui possèdent de nombreux liens et des agents périphériques qui en possèdent peu et seulement avec des agents centraux) ou si au contraire la répartition des liens est plus homogène. De plus, cela complète l'analyse que l'on peut faire de la centralité d'un agent : plus le graphe sera centralisé, plus les agents possédant un indicateur élevé ont un pouvoir ou un rôle important au sein de ce graphe.

Nous présentons l'indicateur dans le cas le plus simple c'est-à-dire celui d'un réseau binaire. L'indicateur se calcule de la manière suivante :

$$CD^B(G) = \frac{\sum_{i=1}^n [CD^B(n^*) - CD^B(i)]}{\max \sum_{i=1}^n [CD^B(n^*) - CD^B(i)]}$$

où $CD^B(n^*)$ est le degré observé le plus élevé dans le réseau étudié alors $\max \sum_{i=1}^n [CD^B(n^*) - CD^B(i)]$ est le maximum théorique que l'on pourrait obtenir dans le réseau étudié.

Si l'on reprend l'exemple 1, l'indicateur est égal à :

$$\begin{aligned} CD^B(G) &= \frac{(5-5)+(5-2)+(5-3)+(5-2)+(5-3)+(5-1)+(5-2)}{(7-1)(7-2)} \\ &= \frac{17}{30} \\ &= 56,67\% \end{aligned}$$

Le degré du réseau correspond à 56,65% du maximum qu'il pourrait atteindre si le réseau de l'exemple 1 était en forme d'étoile et présentait la

¹Un certain nombre de concepts renvoient à la notion d'équivalence au sein du réseau pour deux agents (Lorrain et Harrison, 1971 ; Sailer, 1978 ; Burt, 1983 ; Doreian, 1988 ; Wasserman et Faust, 2007)

forme où un agent est en position centrale. Le réseau de l'exemple 1 est donc assez centralisé.

Remarque : L'adaptation pour un réseau valué est plus délicate dans la mesure où on ne peut pas réellement définir le degré maximum théorique du réseau. Nous pouvons juste faire une hypothèse simplificatrice qui serait de considérer que ce maximum est égal à $(n - 1)$ fois le lien qui possède la plus grande valeur.

Dans le paragraphe suivant, nous présentons un indicateur de centralité qui s'appuie sur la notion de distance.

1.4.1.2. Centralité de proximité ou “closeness centrality”

La centralité de proximité ou “closeness centrality” (ou plus simplement closeness) est une mesure différente du degré. Un agent est considéré comme central s'il est situé à proximité de tous les autres agents du réseau, c'est-à-dire s'il peut rapidement interagir avec les autres agents. Cette définition de la centralité met en avant la capacité d'un agent à pouvoir faire partager rapidement une information dans le réseau et, réciproquement, à pouvoir rapidement capter toutes les informations qui circulent dans le réseau. Le fait de pouvoir rapidement accéder à une information est évidemment un avantage dans un environnement comme celui de la recherche. C'est un gain de temps et cela limite les coûts d'accès à l'information. De plus, si on fait l'hypothèse que la circulation de l'information au sein d'un réseau n'est pas parfaite (réseau borné ou hypothèse de “decay”) alors plus l'agent est proche des différentes sources d'informations plus il est en mesure de capter des informations. Cela s'applique aussi pour les externalités de connaissance.

La centralité de proximité d'un agent i est calculée de la manière suivante. Tout d'abord, on calcule la distance géodésique de l'agent i à tous les autres agents du réseau. Puis, on effectue la somme des distances ainsi obtenues afin d'obtenir le “farness” de l'agent i . Enfin, on calcule l'inverse du “farness” pour obtenir le “closeness” de l'agent i .

La mesure de “closeness” pour un agent est égale à :

$$CC(i) = \frac{1}{\left(\sum_{j \neq i} d_{ij}\right)}$$

avec $i, j \in N$

et la mesure normalisée est égale à :

$$CC'(i) = \frac{n-1}{\sum_{j \neq i} d_{ij}}$$

L'indicateur normalisé est multiplié par $(n-1)$ puisque la valeur maximale que peut prendre le closeness est atteinte quand un agent est connecté directement avec tous les autres agents du réseau, ce qui correspond à une valeur de $(n-1)^{-1}$. Notons que la distance géodésique entre deux agents est la même quelle que soit l'intensité des liens empruntés par le chemin qui relie les deux agents.

Afin d'illustrer la manière de calculer l'indicateur de closeness, nous reprenons l'exemple 1 :

Indicateur	$CC(i)$	$CC(i)'$
Agent 1	7	$\frac{6}{7}$
Agent 2	11	$\frac{6}{11}$
Agent 3	10	$\frac{6}{10}$
Agent 4	10	$\frac{6}{10}$
Agent 5	9	$\frac{6}{9}$
Agent 6	14	$\frac{6}{14}$
Agent 7	11	$\frac{6}{11}$

Tableau 4 : Indicateur de closeness pour les agents de l'exemple 1

Sur ce tableau, l'agent le plus central selon la mesure de closeness est l'agent 1. Les agents 2 et 7 ont des indicateurs de closeness identiques, ce qui est normal dans la mesure où ces deux agents occupent des positions parfaitement identiques au sein du réseau. En revanche, les agents 3 et 4 qui n'occupent pas des positions équivalentes au sein du réseau possèdent le même closeness. Dans ce cas, l'indicateur de closeness ne permet pas de discriminer entre ces deux agents.

Pour calculer le closeness, seule la distance géodésique est utilisée. Or, on peut supposer que les agents, lorsqu'ils cherchent à faire circuler de l'information ou à communiquer au sein d'un réseau, n'utilisent pas uniquement le plus court chemin mais mobilisent tous les chemins possibles. Il existe différentes déclinaisons du closeness, notamment une qui prend en compte l'ensemble des chemins possibles pour relier deux agents au sein du réseau.

La principale difficulté pour calculer l'indicateur de "closeness" se pose lorsque le graphe n'est pas connexe, c'est-à-dire lorsqu'au moins un agent du réseau n'est pas lié (indirectement ou directement) avec tous les autres agents du réseau ou, dit autrement, lorsque le réseau possède au moins deux composantes. Dans ce cas, il est impossible d'évaluer la distance en nombre de liens entre deux agents. Comme pour le calcul du diamètre ou de la distance, la solution généralement retenue est de calculer l'indicateur de "closeness" seulement pour les agents de la composante principale. Cette solution est envisageable à condition que la composante principale contienne un nombre assez élevé d'agents relativement à la seconde plus grande composante. De cette manière, on pourrait considérer avec un risque d'erreur minime que l'agent qui a le closeness le plus faible à l'intérieur de cette composante est vraiment le plus central.

Une autre solution souvent envisagée consiste à faire l'hypothèse que si deux agents ne sont pas connectés directement ou indirectement, alors la distance qui les sépare est égale à l'infini. Pour deux agents i et j n'appartenant pas à la même composante, la distance d_{ij} sera égale à $+\infty$. L'avantage de cette hypothèse est qu'elle prend en compte l'impossibilité pour les deux agents de communiquer au sein du réseau. Si cette hypothèse semble être plus proche de la réalité, elle demande néanmoins d'adopter la mesure du closeness. A partir du moment où le graphe compte plusieurs composantes, l'inverse de la somme des distances géodésiques tendraient alors vers 0 pour tous les agents. Nous proposons l'adaptation suivante :

$$CC_A(i) = \sum_{j \neq i} \frac{1}{d_{ij}}$$

avec $i, j \in N$

Dans ce cas, on a :

$$\lim_{d_{ij} \rightarrow \infty} \frac{1}{d_{ij}} = 0^+$$

Grâce à cet indicateur, on peut calculer le closeness pour des graphes qui

ne sont pas connexes.

B. Cornwell (2005) propose une autre méthode qui permet de calculer le closeness des agents dans un réseau non connexe. Pour cela, il utilise le complément du graphe étudié. Le complément d'un graphe G sera noté G_C . Il est défini par le même ensemble d'agents N et l'ensemble des liens est la différence entre l'ensemble des liens possibles et l'ensemble des liens effectivement réalisé. Ainsi si entre deux agents distincts, il n'existe pas de liens directs dans le réseau alors ce lien est présent dans le complément de ce graphe. Si un graphe est déconnecté alors son complément est connecté. De cette manière, on peut alors facilement calculer un indicateur de closeness. Afin d'obtenir l'agent le plus central dans le graphe de départ, on retranche cette mesure à 1 pour obtenir l'agent le plus central dans le réseau d'origine. La comparaison d'un réseau à un autre est possible grâce à une mesure normalisée comprise entre 0 et 1, mesure qui prend en compte la taille initiale de la composante dans laquelle se trouve l'agent. Plus la mesure est proche de 1, plus l'agent est central dans le complément du graphe.

De la même manière que l'on a calculé un indicateur de degré pour le réseau, on peut calculer un indicateur de closeness. Le calcul de l'indicateur est presque identique. Dans le cas du closeness, nous retenons l'indicateur normalisé. L'indicateur se calcule de la manière suivante :

$$CC(G) = \frac{\sum_{i=1}^n [CC'(n^*) - CC'(i)]}{\max \sum_{i=1}^n [CC'(n^*) - CC'(i)]}$$

où $CC'(n^*)$ est le closeness observé le plus élevé dans le réseau étudié alors $\max \sum_{i=1}^n [CC'(n^*) - CC'(i)]$ est le maximum théorique que l'on pourrait obtenir dans le réseau étudié.

Pour l'exemple 1, nous obtenons le résultat suivant :

$$\begin{aligned} CC(G) &= \frac{\sum_{i=1}^n [CC'(n^*) - CC'(i)]}{(n-1)(1 - \frac{n-1}{2(n-2)+1})} \\ &= \frac{\sum_{i=1}^n [CC'(n^*) - CC'(i)]}{\frac{(n-1)(n-2)}{(2n-3)}} \\ &= 64,41\% \end{aligned}$$

Cet indicateur confirme que le réseau de l'exemple 1 est assez fortement centralisé.

La mesure de "closeness" est plus précise que le degré au sens où la centralité de l'agent n'est pas seulement mesurée de manière locale mais relativement à l'ensemble des agents du réseau. De plus cette mesure est plus

discriminante que le degré. Toutefois, il est toujours possible pour deux agents n'occupant pas une position équivalente au sein du réseau d'obtenir une mesure de closeness identique. Plus particulièrement, on remarque que les agents qui sont situés à une extrémité du réseau et qui ne possèdent qu'un seul lien ont tous un indicateur de closeness qui vaut 0. Or, ces agents n'occupent pourtant pas forcément des positions identiques notamment pour capter les connaissances et les informations qui circulent dans le réseau. De plus, cette mesure apparaît imprécise lorsque le graphe n'est pas connexe. Afin de pouvoir calculer le closeness pour les différents agents du réseau, nous sommes obligés de faire des hypothèses simplificatrices. Enfin, cet indicateur, même s'il prend en compte la position respective des autres agents dans le calcul de la centralité d'un agent, ne prend pas en compte l'importance des agents dont il est le plus proche.

Remarque : La valeur des liens ne peut pas être prise en compte quand on étudie des réseaux issus des PCRDT dans lesquels les flux qui circulent sont majoritairement des actifs intangibles comme des informations, des connaissances ou encore des externalités et que la valeur du lien représente seulement un compte du nombre de collaborations. Une manière de prendre en compte la valeur des liens serait de considérer par exemple seulement les liens d'une valeur supérieure à 1. Puis, recalculer dans ce nouveau réseau l'indicateur de closeness afin d'observer quels sont les agents qui, dans un réseau composé de relations réitérées au sein d'un même programme cadre, sont les plus centraux. Cette difficulté est inhérente à l'ensemble des indicateurs qui utilisent uniquement le concept de distance.

Dans le paragraphe suivant, nous présentons un indicateur qui permet de mesurer différemment la centralité en cherchant à mettre en avant le rôle joué par un agent dans le fonctionnement du réseau et notamment dans sa capacité à être un intermédiaire au sein de celui-ci.

1.4.1.3. Centralité d'intermédiation ou “betweenness centrality”

Cet indicateur met en avant un concept de centralité différent de ceux vus précédemment. Il n'est pas basé sur la notion de liens ou de capacité à créer des partenariats mais sur la capacité d'un agent à se rendre important au sein du réseau en nouant les bons partenariats. Contrairement au closeness, cet indicateur peut être calculé même si le graphe n'est pas connexe.

Un agent est central s'il facilite la circulation des informations et des connaissances entre les différents agents du réseau. Le “betweenness” comptabilise le nombre de fois qu'un agent est situé sur le plus court chemin entre

toutes les autres paires d'agents. Pour déterminer la centralité d'un agent i , on compte le nombre de chemins géodésiques entre deux agents distincts qui passent par cet agent. Puis, on réitère cette mesure pour chaque paire d'agents. Cette statistique permet d'évaluer le rôle joué par un agent dans les relations entre deux autres agents. Un agent situé entre deux autres agents peut empêcher toute communication au sein du réseau. Cette statistique repose sur l'hypothèse que deux agents qui souhaitent communiquer passent par le plus court chemin.

Au sein d'un réseau de collaborations en R&D, un agent qui est nécessaire à la bonne communication au sein du réseau et qui permet aux informations de circuler est un rouage important du réseau. Sa position constitue de plus un avantage non négligeable. Sans cet agent, il est possible que le réseau soit moins efficace. Les agents qui ont une centralité d'intermédiarité élevée sont souvent des agents qui participent à plusieurs projets au sein des programmes cadres et qui servent donc de liens entre différents projets. Si on supprimait ces agents du réseau, on pourrait s'attendre à un impact sur les propriétés générales de celui-ci. Les agents qui possèdent un *betweenness* élevé permettent au réseau de collaborations d'être un réseau intégré et connexe. Par exemple, un réseau composé de deux projets avec un seul agent commun est connexe. Si cet agent est supprimé, le réseau est déconnecté. Du fait de sa position d'intermédiaire cet agent a un fort *betweenness*.

L'indicateur de "betweenness" a pour particularité de se calculer de manière identique quelle que soit la représentation choisie. Il existe une adaptation de cet indicateur au réseau valué mais elle est importante si l'intensité des liens renvoie à des flux quantifiables comme par exemple les réseaux informatiques ou les réseaux routiers lorsque l'on s'intéresse aux effets de saturation ou de goulot d'étranglement. Dans ces cas, on cherche à vérifier les chemins optimaux pour pouvoir faire circuler de l'information ou des véhicules. Dans notre cas, l'intensité des liens renvoie à un nombre de collaborations, collaborations qui contribuent à la circulation des informations et des externalités qui sont non mesurables. C'est pourquoi nous n'utilisons pas cette adaption.

L'indicateur de "betweenness" se calcule de la manière suivante. Soient GP_{jk} le nombre de chemins géodésiques entre l'agent j et l'agent k avec $j, k \in N$ et $GP_{jk}(i)$ le nombre de chemins géodésiques entre j et k qui passent par l'agent i avec $i \in N$. L'indicateur de *betweenness* est alors défini de la manière suivante :

$$CB(i) = \sum_{j < k} \frac{GP_{jk}(i)}{GP_{jk}}$$

La mesure normalisée est la suivante :

$$CB'(i) = \frac{CB(i)}{(n-1)(n-2)/2}$$

Le minimum que peut atteindre cet indicateur est zéro et le maximum est $(n-1)(n-2)/2$ qui est le nombre maximal de paires dans le réseau sans prendre en compte l'agent i .

La mesure de centralité qui permet de comparer le réseau avec un réseau en forme d'étoile est la suivante :

$$CB(G) = \frac{\sum_{i=1}^n [CB(n^*) - CB(i)]}{\max \sum_{i=1}^n [CB(n^*) - CB(i)]}$$

ce qui se simplifie de la manière suivante pour le betweenness :

$$CB(G) = \frac{\sum_{i=1}^n [CB(n^*) - CB(i)]}{(n-1)^2(n-2)}$$

On peut calculer cet indicateur à partir des mesures normalisées.

$$CB(G) = \frac{\sum_{i=1}^n [CB'(n^*) - CB'(i)]}{(n-1)}$$

Pour illustrer cet indicateur, nous reprenons l'exemple 1.

Indicateur	$CC(i)$	$CC'(i)$
Agent 1	9	63,33
Agent 2	0	0
Agent 3	0,5	3,33
Agent 4	0	0
Agent 5	5	33,33
Agent 6	0	0
Agent 7	0	0

Tableau 5 : Indicateur de “closeness” pour les agents de l’exemple 1

Sur l’exemple ci-dessus, on voit clairement que ce sont les agents qui participent à plusieurs projets qui sont les plus centraux. Ce sont même les seuls qui ont un betweenness strictement supérieur à 0. Cette mesure permet de mettre en avant les agents qui servent à faire la liaison entre les différents projets au sein du réseau et seulement ceux là. Puisque la thèse portent sur deux domaines technologiques proches, il est intéressant de vérifier si, en agrégeant les deux réseaux, les agents présents dans les deux domaines technologiques sont ceux qui ont le betweenness le plus élevé comme Breschi et Catalini l’ont mis en évidence dans leur article de 2007.

L’indicateur de “betweenness” permet d’extraire les agents du réseau qui ont un indicateur de degré élevé mais qui ne participent qu’à un seul projet. En effet, par définition du betweenness, un agent qui ne participe qu’à un seul projet a un indicateur égal à 0. De fait, il est possible, en utilisant le betweenness, de repérer les agents qui ne participent qu’à un seul projet.

Lorsque les agents cherchent à communiquer ou lorsque l’on cherche à déterminer s’ils occupent une place centrale (au sens du betweenness) dans le réseau, on suppose qu’ils utilisent uniquement les chemins géodésiques pour communiquer. Or, on peut supposer si un agent souhaite communiquer une information à un autre agent ou a besoin d’obtenir une information, il peut mobiliser différents liens afin d’atteindre son objectif via différents chemins et pas uniquement par les chemins géodésiques. Pour remédier à cette difficulté un certain nombre de chercheurs ont créé des mesures alternatives qui prennent en compte tous les chemins possibles dans le calcul du betweenness.

Enfin, s’il est intéressant de mettre en avant seulement les agents qui présentent des caractéristiques particulières, notamment ceux qui participent à plusieurs projets, on peut avoir besoin de discriminer parmi les autres agents du réseau. En effet, si on observe les mesures obtenues à l’exemple 1, on s’aperçoit que 4 agents sur 7 (les agents 2, 4, 6 et 7) ont la même centralité. Pour autant, ces 4 agents n’occupent pas une place similaire au sein du réseau (à part les agents 2 et 7). Mis à part l’agent 6, qui est situé à une extrémité du graphe et ne peut donc pas servir d’intermédiaire, les trois autres agents peuvent être des intermédiaires même s’ils ne sont pas situés sur des chemins géodésiques. Il apparaît donc que le betweenness ne permet pas de discriminer entre certains agents qui occupent des positions différentes au sein du graphe.

Dans le paragraphe suivant, nous présentons l’eigenvector centrality. Cet

indicateur contrairement au betweenness présente comme caractéristique d'être fortement discriminant.

1.4.1.4. Centralité par les valeurs propres ou "eigenvector centrality"

L'approche par les valeurs propres cherche à mesurer la centralité des agents du réseau en prenant en compte la centralité des autres agents présents dans le réseau. La centralité par les valeurs propres suppose que les connaissances, les informations et les externalités de connaissance peuvent être véhiculées au travers de n'importe quel lien et ne sont pas contraintes de passer seulement par les chemins géodésiques. Plus précisément, pour un agent i , sa centralité dépend des agents avec qui il est directement lié et, au travers de la centralité de ces partenaires, dépend au final de la centralité de l'ensemble des agents avec qui il est indirectement lié (Bonacich 1972a, 1978, 1991).

Cette mesure est basée sur l'idée d'influence indirecte. Ainsi, même si un agent n'influence qu'un seul autre agent, si ce dernier influence de nombreux autres agents (qui eux mêmes font la même chose) alors le premier agent de la chaîne possède une certaine influence. Dans un réseau de collaborations, il est évident qu'il est important d'être situé à proximité, si possible d'être en contact direct, avec les agents les plus influents du réseau.

La centralité par les valeurs propres permet de mesurer la centralité d'un agent en prenant en compte à la fois sa position dans le réseau et la centralité de ses partenaires. La centralité d'un agent est proportionnelle à la centralité des agents avec qui il est lié et à la force de ses liens. La centralité par les valeurs propres s'adapte parfaitement aux cas des PCRDT puisqu'il est possible de calculer cet indicateur aussi bien pour le réseau binaire que valué.

Nous présentons cet indicateur dans le cas du réseau binaire. Soit $CE(i)$ la centralité de l'agent i et x_{ij}^B l'élément de la matrice X^B qui rend compte de la relation entre l'agent i et l'agent j . La centralité de l'agent i est définie de la manière suivante :

$$CE^B(i) \approx \sum_{j \in N \setminus i} CE(j) \times x_{ij}^B$$

pour $i, j \in N$.

Afin de calculer la centralité d'un agent à l'intérieur d'un réseau, nous devons résoudre un système d'équations linéaires simultanées. Ce système

peut être exprimé comme un problème de valeurs propres et de vecteurs propres associés aux valeurs propres. La solution est donnée par :

$$X^B.CE^B = \lambda CE^B$$

avec X^B la matrice d'adjacence, λ la plus grande valeur propre (selon le théorème de Perron-Frobenius, cette valeur permet d'obtenir un vecteur propre associé à la valeur propre la plus élevée qui contient uniquement des entrées positives) et CE^B le vecteur propre associé à la valeur propre λ qui donne l'ensemble des indicateurs de centralité. Par la suite, pour pouvoir réaliser des comparaisons d'un réseau à un autre, la normalisation retenue est que la somme des indicateurs normalisés au carré doit être égale à 1.

Afin d'illustrer la manière de calculer cette mesure, nous reprenons l'exemple 1 :

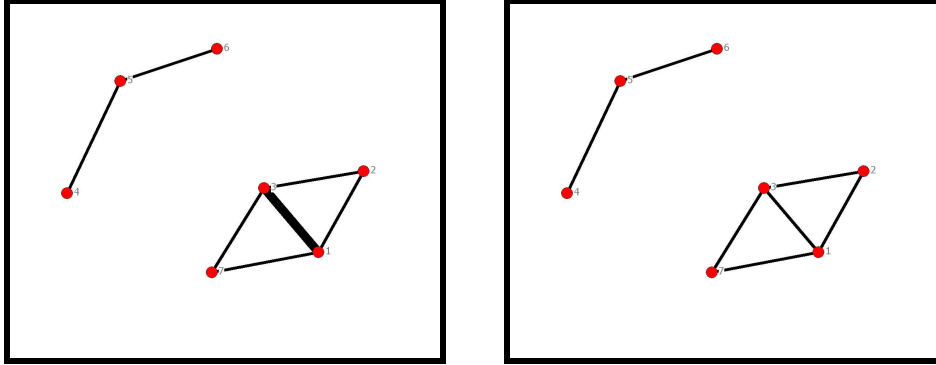
Indicateur	$CE^V(i)$	$CE^V(i)'$	$CE^B(i)$	$CE^B(i)'$
Agent 1	0,710	100,392	0,715	101,147
Agent 2	0,254	35,966	0,270	38,130
Agent 3	0,509	71,933	0,400	56,566
Agent 4	0,201	28,460	0,254	35,901
Agent 5	0,254	35,966	0,335	47,348
Agent 6	0,053	7,507	0,081	11,447
Agent 7	0,254	35,966	0,270	38,130

Tableau 6 : Indicateur d'“eigenvector centrality” pour les agents de l'exemple 1

Le principal avantage de cet indicateur est qu'il est facilement transposable d'un réseau à un autre. Il mesure la centralité d'un agent en prenant en compte la position de l'agent dans le réseau de manière globale puisque la centralité de cet agent dépend des agents avec qui il est lié directement et indirectement.

L'indicateur d'"eigenvector centrality" malgré ces limites est un indicateur de référence. Il est particulièrement intéressant dans le cadre des réseaux de collaborations en innovation. En effet, il permet de prendre en compte dans la centralité d'un agent celle de ces partenaires et, de fait, de manière indirecte, celle des partenaires de ces partenaires et ainsi de suite. Au final, la centralité d'un agent dépend de celle de l'ensemble des agents qui appartiennent à la même composante que lui. Puisqu'il existe au moins un lien indirect entre tous ces agents, on peut supposer que les informations, les connaissances et les externalités circulent au sein de la composante. Pour les agents en dehors de cette composante, il est impossible de tirer profit de ces flux immatériels du fait de l'absence totale de contact avec au moins un des membres de cette composante. De plus, si la centralité d'un agent dépend de la centralité des agents de la composante, cet indicateur prend aussi en compte l'intensité des liens. De cette manière, il est possible de tenir compte de l'intensité de certains partenariats au sein de la composante. Enfin, comme la centralité d'un agent dépend directement de celle de ses partenaires et indirectement de celle des partenaires des partenaires, cet indicateur respecte l'hypothèse de "decay".

L'indicateur de centralité par les valeurs propres possède cependant deux limites. La première est qu'il est très difficile à calculer lorsque le nombre d'agents devient très grand étant donné qu'il s'agit d'un système de n équations à n inconnues. La seconde plus problématique est le fait que cet indicateur n'est calculé que pour la grande composante. En effet, si le réseau contient plusieurs composantes, cela revient à résoudre un système d'équations pour chaque composante où le nombre d'équations et d'inconnus sont identiques au nombre d'agents au sein de la composante. Nous obtenons alors un indicateur pour chaque agent de chaque composante. Cet indicateur permet bien de classer chaque agent d'une composante du plus central au moins central mais ne permet pas de classer les agents entre deux composantes. La difficulté consiste à trouver une échelle de valeur qui permettrait de prendre en compte les différences de taille des composantes. Cette difficulté est illustrée par l'exemple 2 ci-dessous.



Graphes 9 et 10 : Exemple de graphes non connexes

Sur ce graphe, on distingue deux composantes. Si l'on prend le cas où le graphe est valué, on a la matrice d'adjacence suivante :

Agents	Ag. 1	Ag. 2	Ag. 3	Ag. 4	Ag. 5	Ag. 6	Ag. 7
Agent 1	0	1	2	0	0	0	1
Agent 2	1	0	1	0	0	0	0
Agent 3	2	1	0	0	0	0	1
Agent 4	0	0	0	0	1	0	0
Agent 5	0	0	0	1	0	1	0
Agent 6	0	0	0	0	1	0	0
Agent 7	1	0	1	0	0	0	0

$Ag.^* = Agent$

A partir de cette matrice X^V , le système de 7 équations à 7 inconnus que l'on cherche à résoudre est le suivant :

$$X^V.CE^V = \lambda.CE^V$$

CE^V est le vecteur propre associé à la valeur propre λ . Ce système peut s'écrire de la manière suivante :

$$\begin{pmatrix} 0 & 1 & 2 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix} \times \begin{pmatrix} CE^V(1) \\ CE^V(2) \\ CE^V(3) \\ CE^V(4) \\ CE^V(5) \\ CE^V(6) \\ CE^V(7) \end{pmatrix} = \lambda \times \begin{pmatrix} CE^V(1) \\ CE^V(2) \\ CE^V(3) \\ CE^V(4) \\ CE^V(5) \\ CE^V(6) \\ CE^V(7) \end{pmatrix}$$

Nous observons que la centralité d'un agent au sein du graphe ne dépend que des agents qui sont dans la même composante que lui. Par exemple, la centralité de l'agent 1 dépend de celle de l'agent 2, de l'agent 3 et de l'agent 7 mais ne dépend pas des agents 4, 5 et 6. Le logiciel Ucinet qui est utilisé pour les calculs d'indicateurs ne renvoie des valeurs que pour la plus grande composante.

Par conséquent dans cet exemple, la centralité des agents 1, 2, 3 et 7 est liée et celle des agents 4, 5 et 6 aussi. Il est alors difficile de déterminer la centralité d'agents appartenant à des composantes différentes car se pose alors la question d'évaluer l'importance de chaque composante.

Calculer exclusivement la centralité des agents de la grande composante est une solution qui peut être retenue si la grande composante du graphe contient un nombre d'agents que l'on juge "suffisant" ou si la différence de taille entre la plus grande composante et la seconde est "suffisante". Par "suffisant", on sous-entend qu'il faut pouvoir être certain que la plus grande composante contient l'agent le plus central. Dans le cas des PCRDT, il est courant d'observer des composantes qui contiennent plus de 90% des agents du réseau (cf. section 2). Il est alors moins gênant de ne calculer la centralité que pour ces agents.

Tout au long de cette sous-section, nous avons présenté les indicateurs les plus usités pour déterminer les agents centraux d'un réseau. Toutefois, ces indicateurs présentent certaines limites notamment lorsque le graphe n'est pas connexe. Dans la sous-section suivante, nous mettons en avant trois indicateurs qui semblent pouvoir résoudre une partie des difficultés rencontrées par les quatre premiers présentés ou qui permettent d'apporter un surplus d'information pour l'analyse.

1.4.2 Nouvelle approche de la centralité

Dans cette sous-section, nous présentons trois nouveaux indicateurs dont deux ont été produits durant cette thèse. Ces indicateurs permettent de contourner un certain nombre de difficultés rencontrées par les indicateurs de centralité présentés précédemment.

1.4.2.1. Le degré renforcé

Nous proposons un autre indicateur afin d'améliorer l'analyse notamment dans le cadre des réseaux non connexes. Cet indicateur est basé sur le degré qui est une mesure simple. Nous avons vu que le degré d'un agent indique le nombre de liens que possède l'agent au sein du graphe. Cependant, cette

mesure ne prend pas en compte la centralité des autres agents avec qui l'agent est lié comme le fait la centralité par les valeurs propres. Cette dernière mesure est toutefois délicate à appliquer lorsque le graphe du réseau n'est pas connexe et, encore plus, lorsqu'aucune composante du graphe n'est de taille "suffisante".

Notre indicateur essaie de contourner les difficultés rencontrées par l'indicateur de centralité par les valeurs propres. Plus précisément, l'objectif de notre indicateur est de calculer la centralité d'un agent en prenant en compte son degré et le degré des autres agents du réseau. Nous cherchons à construire un indicateur qui puisse être applicable à toute forme de réseau notamment aux réseaux non connexes. Notre indicateur est basé sur le degré et la distance géodésique entre les agents. La centralité d'un agent est notée $C_{SD}(i)$.

Nous pouvons donner l'interprétation suivante à cet indicateur. Si l'on se réfère à la théorie de la gravitation, l'attraction d'un organisme distant est supposée être égale à sa masse pondérée par une fonction décroissante de la distance. Par conséquent, nous considérons les agents du réseau à atteindre comme des attracteurs et la distance comme une fonction d'impédance (friction de la distance). Plusieurs indicateurs d'accessibilité ont été élaborés par des géographes (Vickerman et al., 1999). Ils ont été utilisés par des économistes afin de mesurer des potentiels de marché et/ou l'accessibilité à la connaissance dans un cadre géographique avec une simple fonction inverse pour modéliser la perte d'information au fur et à mesure de l'accroissement de la distance. L'idée est de considérer que la centralité d'un agent n'est pas liée seulement à sa propre activité (son degré) mais elle dépend aussi de sa capacité à accéder à l'activité menée par tous les autres agents à l'intérieur du réseau.

La mesure la plus simple de la distance dans un réseau relationnel comme le sont les réseaux de collaborations des PCRDT est la distance géodésique entre deux agents et cette distance est considérée comme infini s'il n'existe aucun chemin entre deux agents. Afin de calculer le degré renforcé, nous utilisons le vecteur qui renseigne la centralité de degré de l'ensemble des agents du réseau C_D et la matrice des distances géodésiques qui renvoie l'ensemble des distances géodésiques entre toutes les paires d'agents notées GD . Nous notons gd_{ij} un élément de GD qui est la distance géodésique entre l'agent i et l'agent j , i et j appartenant à N .

La centralité d'un agent i dépend de la centralité de tous les autres agents du réseau mais de manière inversement proportionnelle à la distance qui le sépare de ces agents. Avec cet indicateur, un agent est central s'il est connecté à d'autres agents centraux (selon la mesure de degré) ou s'il est localisé près d'autres agents centraux même s'il est seulement indirectement connecté à

ces derniers.

Afin de calculer le degré de centralité renforcé, nous retenons l'hypothèse suivante : la distance entre deux agents non liés directement est égale à l'infini. De plus, pour un agent i quelconque, le degré de l'agent i est pris en compte. Pour pouvoir utiliser le degré de l'agent i , il faut modifier la matrice de distance géodésique. La diagonale de GD est entièrement composée de zéros. Lorsque l'on souhaite calculer le degré renforcé de l'agent i , nous devons pouvoir prendre en compte sa propre activité donc son propre degré. De ce fait, nous rajoutons un à tous les éléments de la matrice GD de manière à obtenir des 1 sur la diagonale et conserver la notion de distance. Ensuite, nous prenons l'inverse de chaque élément de GD . De cette façon, si la distance entre deux agents est égale à l'infini, alors l'élément de GD^* associé à ces deux agents sera égal à zéro. De cette manière, l'activité d'un agent dépend uniquement de l'activité des agents situés dans sa composante et de la sienne, ce qui permet de prendre en compte de manière indirecte la taille de la composante à laquelle appartient un agent dans le calcul de son degré renforcé. Notons GD^* la matrice transformée obtenue à partir de GD et gd_{ij}^* un élément de cette matrice transformée. Un élément gd_{ij}^* de GD^* s'obtient de la manière suivante :

$$gd_{ij}^* = \frac{1}{gd_{ij} + 1}$$

En utilisant une fonction inverse de manière à prendre en compte la décroissance de l'importance d'un agent selon la distance², le vecteur de centralité de degré renforcé C_{SD} est le suivant :

$$C_{SD} = G^* . C_D \text{ avec } g_{ij}^* = \frac{1}{1 + g_{ij}}$$

La centralité de degré renforcé d'un agent i est :

$$C_{SD}(i) = \sum_{j \in N} g_{ij}^* \times C_D(j)$$

Si on applique cet indicateur à l'exemple 1, nous obtenons les résultats suivants :

²Pour une discussion concernant la pertinence de cette forme voir le papier de Kwan (1998).

	Réseau valué	Réseau binaire
Agent 1	12,88	11,33
Agent 2	9,58	8,58
Agent 3	10,91	9,41
Agent 4	9,5	8,66
Agent 5	10,16	9,33
Agent 6	7,16	6,58
Agent 7	9,58	8,58

Tableau 7 : Indicateur de degré renforcé pour les agents de l'exemple 1

Il existe deux possibilités pour normaliser cet indicateur. La première consiste à rechercher quand un agent peut obtenir le degré renforcé maximum. Ce cas se produit lorsque le graphe est complet. Le degré renforcé maximum pour un agent est égal à $n(n - 1)$. Pour obtenir un indicateur normalisé, il faut alors diviser par ce nombre. Le problème est que dans le réseau valué un agent peut avoir plus de $n - 1$ liens puisqu'il peut être lié plusieurs fois au même agent.

Une autre possibilité consiste à diviser le degré renforcé par le nombre d'agents du graphe, c'est-à-dire n . De cette manière, l'indicateur normalisé de l'agent i peut être interprété comme un niveau de ressource moyen que l'agent i puise dans chaque agent du réseau. Plus un agent arrive à puiser des ressources des autres agents plus il est central. Le problème est que cette deuxième possibilité ne prend pas en compte réellement la taille des réseaux. En effet, entre deux réseaux de tailles différentes mais de densités identiques, le nombre de liens de chaque réseau compte et rend la comparaison entre deux réseaux délicate.

Le degré renforcé permet une meilleure discrimination entre les agents. En effet, une des limites du degré provient du fait que des agents aux positions différentes au sein du réseau peuvent avoir le même indicateur, celui-ci ne prenant en compte qu'une dimension locale. Le degré renforcé nous permet d'ordonner les agents de manière plus précise en termes de centralité. Ainsi, si deux agents possèdent le même degré renforcé, alors ils occupent une position similaire au sein du réseau (sauf cas exceptionnel).

L'indicateur de degré renforcé a l'avantage de pouvoir supporter différentes hypothèses afin de prendre en compte l'environnement propre à chaque réseau. Il est possible par exemple de renforcer l'aspect perte d'influence en fonction de la distance en utilisant des formes quadratiques, on aurait alors les éléments de $g_{ij}^* = 1/(g_{ij} + 1)^2$. Il est aussi possible d'introduire des hypothèses différentes comme de limiter la distance d'influence comme cela existe

dans des modèles théoriques dits bornés (Jackson et Wolinski, 1996) ou encore d'introduire un seuil à partir duquel un agent n'aurait pas d'influence sur un autre. La centralité d'un agent ne dépendrait alors plus que des agents situés à une distance inférieure à ce seuil. Si nous étudions les externalités de connaissance par exemple, nous pouvons supposer qu'un agent ne peut profiter des connaissances d'un autre agent qu'en deçà d'une certaine distance. Enfin, il est possible de combiner à la fois un réseau borné et un réseau où l'influence serait décroissante avec la distance.

Par définition, l'indicateur de degré renforcé est basé sur la distance géodésique et sur le degré. Il est possible d'envisager de prendre en compte l'ensemble des chemins et pas seulement les chemins géodésiques. De même, il peut être intéressant selon l'objet de l'étude d'utiliser d'autres indicateurs de centralité que le degré. Le *betweenness* pourrait par exemple être utilisé si l'on étudiait des réseaux de transport dans lesquels les agents importants sont les agents intermédiaires.

L'indicateur de degré renforcé se rapproche de celui de Katz (1953) et de l'indicateur d'accessibilité de Doreian (1974). Ces deux indicateurs sont proches au niveau de l'intuition de l'eigenvector centrality proposé par Bonacich (1981).

Le premier est basé sur la notion d'influence. Ainsi, un agent est d'autant plus central qu'il peut influencer un nombre important d'agents et son influence est d'autant plus grande qu'il est situé à proximité de ces agents. De la même manière, cet indicateur utilise un paramètre d'atténuation qui peut être plus ou moins important (il est compris entre 0 et 1, 0 étant une totale atténuation et 1 une absence d'atténuation pondérée aussi en fonction de la distance). A la différence de notre indicateur, celui de Katz prend en compte tous les chemins possibles. Contrairement à la définition du chemin que nous avons donné, cet indicateur prend aussi en compte les chemins où le même agent et/ou le même lien apparaît plusieurs fois. Ceci signifie que deux agents à distance 1 sont aussi à distance 3, 5, 7, ... Il existe une certaine redondance dans la prise en compte de ces chemins car un agent ne mobilisera pas plusieurs fois le même lien pour passer une information ou atteindre un agent. Notons que cet indicateur n'utilise pas le poids de l'agent dans le réseau, en d'autres termes chaque agent pèse un poids identique. Or, il peut être pertinent d'intégrer ces différences car il est plus intéressant de se trouver à proximité d'agents plus actifs.

L'indicateur d'accessibilité (Doreian, 1974) est plus proche encore du degré renforcé. Tout comme notre indicateur, il est basé sur la notion de distance géodésique et utilise une pondération. La principale différence est qu'il n'accorde aucun poids aux agents. Tous les agents possèdent la même valeur.

Ainsi ce qui compte pour calculer la centralité d'un agent c'est le nombre distinct d'agents qu'il peut atteindre pour chaque distance géodésique. Prendre en compte l'activité des agents au travers du degré et intégrer cette mesure dans la centralité des agents apparaît être une meilleure solution dans la mesure où il n'est pas équivalent d'être directement lié à l'agent qui compte le plus de liens que d'être lié à un agent qui ne possède qu'un seul lien.

Par rapport à l'indicateur de Bonacich qui utilise une approche par les valeurs propres, notre indicateur apparaît moins performant puisque nous définissons dès le départ le poids de chaque agent par l'intermédiaire du degré pour obtenir par la suite un indicateur de centralité renforcé. Cela revient donc à faire l'hypothèse que le degré est pertinent pour mesurer le poids de chaque agent. Dans notre cas, on mesure l'activité de l'agent grâce à sa capacité à créer des liens. Toutefois, notre indicateur présente l'avantage de pouvoir être calculé même quand le graphe n'est pas connexe et ce pour tous les agents du réseau quelle que soit la composante à laquelle ils appartiennent. De plus, cet indicateur est plus facilement calculable.

Précisons toutefois que d'autres mesures peuvent être envisagées pour remplacer le degré. On peut aussi aborder la notion de poids de chaque agent non pas sous l'angle d'un indicateur réseau mais d'une variable que l'on possède sur chaque agent qui rendrait compte de l'importance de l'agent en R&D. Cette variable peut être issue des informations que l'on possède dans la base PCRDT ou des informations extérieures aux PCRDT. Par exemple, chaque agent peut se voir attribuer un poids qui serait le nombre de chercheurs qu'il possède ou le nombre de brevets déposés dans les dernières années. Le degré renforcé est utile par sa flexibilité.

Dans la sous-section suivante, nous présentons des indicateurs qui ont pour objet de mesurer l'impact d'un agent sur l'efficacité du réseau en matière de circulation d'informations et de connaissances.

1.4.2.2. Mesurer l'impact d'un agent sur la connectivité du réseau

Au sein des réseaux des PCRDT, il est difficile d'établir réellement quels sont les flux qui circulent entre agents, leur nature exacte et de les quantifier. Étant donné l'objet et les objectifs des PCRDT, nous pouvons supposer que circulent au sein des réseaux, des connaissances et des informations susceptibles d'améliorer l'efficacité des agents présents au sein du réseau en matière d'innovation.

Les indicateurs proposés dans cette sous-section cherchent à mesurer

quel serait la perte d'efficacité dans la transmission des informations et des connaissances si un agent ne faisait plus partie du réseau. Nous avons vu à la section 1.3 que la distance moyenne, le diamètre ainsi que le nombre de composantes d'un réseau sont autant d'indicateurs qui permettent d'évaluer l'efficacité du réseau. L'objectif est ici de présenter des indicateurs qui prennent en compte ces dimensions pour évaluer la centralité d'un agent. Un agent sera central s'il a un impact sur l'efficacité du réseau. Cette définition de la centralité est différente de celles précédemment évoquées. En effet, un agent qui a de nombreux liens ne sera pas forcément central selon cette définition. La définition présentée ici se rapproche plus de la définition du *betweenness*. Ainsi un agent avec un faible nombre de liens peut très bien être central s'il a la capacité d'empêcher plusieurs agents de communiquer et s'il joue un rôle d'intermédiaire. Ce concept de centralité est d'autant plus important dans les réseaux où l'on suppose que les hypothèses de "decay" et/ou de réseau borné s'appliquent comme c'est le cas dans les réseaux d'innovation (notamment ceux des PCRD).

Pour évaluer l'impact d'un agent il existe plusieurs possibilités que l'on présente ci-après.

“Cut-points” et fragmentation du réseau La première possibilité et la plus simple est d'observer si le fait de supprimer un agent du réseau conduit à éclater le graphe en un nombre de composantes plus élevé qu'au départ. Si un agent joue un rôle intermédiaire entre deux projets et que cet agent est le seul moyen pour les agents des deux projets de communiquer sur le réseau alors cet agent est un “cut-point”, c'est-à-dire un agent qui, s'il est supprimé du réseau, conduit à une fragmentation du réseau. Tous les “cut-points” n'ont pas la même importance selon le nombre de nouvelles composantes créées et la taille de ces composantes.

Borgatti (2006) propose un indicateur qui rend compte de l'importance d'un agent dans le réseau. Cet indicateur se calcule à partir du nombre de composantes du graphe et du nombre d'agents dans chaque composante. Nous présentons cet indicateur (F) dans le cas simple où le graphe de départ est connexe. L'indicateur s'obtient de la manière suivante :

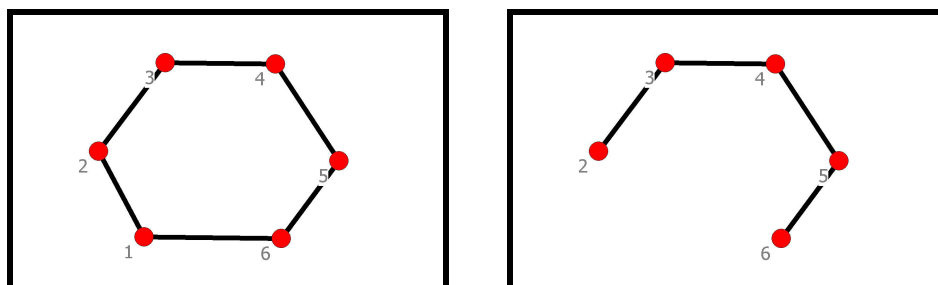
$$F = 1 - \frac{\sum_{k=1}^m n_k(n_k - 1)}{n(n - 1)}$$

avec m le nombre de composantes obtenues après avoir retiré un agent, n_k le nombre d'agents dans la composante k et n le nombre d'agents dans le

graphe de départ. Par construction, cette mesure est comprise entre 0 et 1. Plus elle est proche de 1, plus la suppression de l'agent a un impact sur le graphe.

Lorsque l'on reprend l'exemple 1, on observe que deux agents sont des "cut-points", il s'agit des agents 1 et 5. la mesure de fragmentation pour ces deux agents est respectivement de 0,714 et de 0,524. L'agent 1 a donc un impact plus important sur le graphe. La suppression de chaque agent entraîne la création de deux composantes mais la suppression de l'agent 1 conduit à deux composantes de tailles identiques alors que celle de l'agent 5 produit un singleton et laisse une composante majeure.

L'utilisation de cet indicateur dans le cadre des PCRDT est pertinente dans la mesure où cet indicateur met en avant les agents qui servent de relais. Toutefois, il est possible que l'impact d'un agent sur l'efficacité du réseau soit importante sans pour autant conduire à la création de deux composantes. Par exemple, si l'on observe les graphes ci-dessous, la suppression de l'agent 1 ne crée aucune nouvelle composante mais a un impact sur la distance moyenne et le diamètre du graphe.



Graphes 11 et 12 : Représentation de la suppression d'un agent dans un "one-mode network".

Les réseaux des PCRDT sont constitués d'un grand nombre d'agents qui collaborent au sein de nombreux projets. Il est donc probable qu'il existe plusieurs chemins qui permettront de connecter deux agents et que la suppression d'un agent ne conduira pas nécessairement à la création de nouvelles composantes. Par contre, la suppression de certains agents peut entraîner une modification des propriétés générales du graphe autres que le nombre de composantes. Afin de prendre en considération ces cas de figure, il est donc opportun de combiner cet indicateur à une mesure de l'impact des agents sur les distances entre les agents au sein du réseau. C'est l'objet du paragraphe suivant.

Mesure d'efficience du réseau Pour compléter la mesure de fragmentation, il est intéressant d'utiliser l'indicateur d'efficience proposée par Latora et Marchiori (2007). Cet indicateur est basé sur la notion d'efficience en essayant de déterminer de quelle manière répond le réseau après suppression d'un agent. Pour cela, les auteurs utilisent la notion de distance géodésique combinée à une notion de proximité (type distance géographique par exemple). Latora et Marchiori ont défini l'efficience dans deux articles de 2001 et 2003. L'efficience dans la communication entre l'agent i et l'agent j , notée ε_{ij} , est définie comme le chemin le plus court p_{ij} entre les deux agents. La distance la plus courte ne correspond pas à la distance géodésique puisqu'elle est calculée à partir de la matrice d'adjacence mais aussi à partir de la matrice qui renvoie à la proximité des agents. Cet indicateur est difficile à mettre en œuvre étant donné la combinaison des deux matrices, difficulté croissante avec le nombre d'agents.

L'indicateur d'efficience d'un réseau se définit de la manière suivante :

$$E(g) = \frac{\sum_{i \neq j, j \in N} \varepsilon_{ij}}{n(n-1)} = \frac{1}{n(n-1)} \sum_{i \neq j, j \in N} \frac{1}{p_{ij}}$$

Dans l'équation ci-dessus, on observe que l'on prend l'inverse de p_{ij} de manière à ce que plus les agents sont proches plus l'indicateur est élevé. De plus, dans le cas où deux agents ne sont pas liés dans le réseau, les auteurs font l'hypothèse que $p_{ij} = +\infty$.

À partir de cette mesure, il reste à définir la centralité d'un agent à partir de l'efficience du graphe. La mesure pour un agent i est la suivante :

$$C_i^I = \frac{\Delta E}{E} = \frac{E(g) - E(g'_i)}{E(g)}$$

avec $E(g'_i)$ qui est la mesure d'efficience dans le réseau g dans lequel tous les liens incidents à l'agent i ont été supprimés. Par construction, cette mesure appartient à l'intervalle $[0; 1]$. La suppression d'un agent n'affecte de fait que les agents de sa propre composante. Il est possible d'étendre la mesure à un ensemble d'agents. Ainsi, au lieu de considérer l'impact d'un agent i sur l'efficacité du réseau, on considère l'impact d'un sous-ensemble d'agents noté S . L'indicateur se construit de manière identique. Il est difficile de calculer cet indicateur pour les 2^n sous-ensembles possibles d'agents. Toutefois, cette approche peut être intéressante pour observer l'impact d'un sous-ensemble d'agents que l'on suppose posséder des priorités particulières et que l'on a

identifié au préalable (cf. section 1.5 sur la méthode cœur/périphérie).

Remarque : Il est possible d'adapter cet indicateur en remplaçant l'efficacité définie par Latora et Marchiori par la distance géodésique.

Dans le cas des PCRDT, la matrice évaluée renvoie au nombre de collaborations entre agents. Il est impossible d'utiliser cette matrice comme matrice de proximité dans le cadre de cet indicateur comme on pourrait le faire avec une matrice renvoyant à la distance géographique entre individus par exemple.

Perte de partenariats L'autre manière d'évaluer la centralité d'un agent en termes d'impact sur le réseau est de se placer du point de vue de ses collaborateurs. Ainsi, si l'on supprime un agent du réseau, c'est un collaborateur en moins pour tous les agents qui étaient liés avec lui et un rouage en moins pour faire circuler de l'information pour les agents liés indirectement avec cet agent. Dans tous les cas, la suppression de ce dernier correspond à une perte de connaissances et d'efficacité pour le réseau qui impacte plus ou moins les autres agents selon la position relative de ces agents par rapport à l'agent supprimé. La perte liée à la suppression de l'agent affecte différemment chaque agent. Nous pouvons supposer que plus un agent se situe à proximité de l'agent supprimé plus la perte est importante et d'autant plus importante qu'il n'existe pas d'autres agents pouvant occuper un rôle similaire. L'indicateur que nous proposons ici est complémentaire à l'indicateur précédent car il est basé sur une autre conception de la centralité.

La première étape consiste donc à évaluer la perte d'efficacité pour un agent j au sein du réseau suite à la suppression de l'agent i . L'agent j ne sera affecté que s'il est situé dans la même composante que l'agent i . Il existe différents moyens d'analyser cette perte. Dans un premier temps, nous exprimons cette perte en termes de partenariats. Par exemple, si l'agent j collabore avec l'agent i et qu'au total l'agent j possède n_j liens dans le réseau, alors on peut estimer que la perte entraînée par la disparition de l'agent i est $1/n_j$. De même, si l'agent i et l'agent j sont à distance 2 (en utilisant ici la distance géodésique) alors, nous déterminons combien d'agents distincts l'agent j atteint à distance 2 exactement (les agents à distance 1 ne sont pas compris) n_{j_2} . L'agent i représentera alors $1/n_{j_2}$ des agents que peut atteindre l'agent j à distance 2. La perte de centralité pour l'agent j suite à la suppression de l'agent i se note $CP_i(j)$ et s'obtient de la manière suivante :

$$CP_i(j) = 1/n_{j_{dij}}$$

avec d_{ij} la distance géodésique entre i et j et $n_{j_{dij}}$ le nombre d'agents que peut atteindre l'agent j pour cette distance géodésique. Ce ratio peut s'apparenter à un score qu'obtient l'agent i au vu de sa position relativement à l'agent j . Ainsi si l'agent j a pour seul partenaire l'agent i , le lien entre les deux est donc important pour j car c'est ce lien qui lui permet de se rattacher au reste du graphe. Pour compléter cet indicateur, il faut prendre en compte la distance qui sépare les deux agents. En effet, la perte de centralité n'est pas la même selon que l'agent est à distance 1 ou à distance 2. Au final, le score total obtenu par l'agent i , noté CP_i , est la somme des pertes qu'entraînerait sa suppression pour tous les autres agents du réseau.

$$CP_i = \sum_{j \in N \setminus i} \frac{1}{d_{ij}} \times CP_i(j)$$

Il est impossible d'utiliser ici l'information que l'on possède sur l'intensité des liens. Si la valeur des liens renvoyait à des flux clairement identifiés, il serait alors possible d'adapter l'indicateur ci-dessus. L'autre difficulté est que cet indicateur ne souligne pas assez le rôle joué par les "cut-points" et, de fait, s'applique en complément de la précédente mesure notamment lorsque l'on observe qu'il existe peu de "cut-points".

Pour l'améliorer, il faudrait prendre en compte le changement introduit par la suppression d'un agent au delà de sa relation avec lui. Ainsi, si suite à la suppression d'un agent i , l'agent j perdait plus de la moitié de ses contacts indirects, il faudrait pouvoir le prendre en compte. Seulement, cela demanderait de vérifier pour chaque distance géodésique le nombre de partenaires que pouvait atteindre l'agent j avant et après suppression de l'agent i .

Dans les trois paragraphes précédents, nous avons présenté la mesure de fragmentation de Borgatti (2006) et la mesure d'efficacité de Latora et Marchioni (2007) puis nous avons proposé une mesure basée sur les partenariats. Ces deux dernières mesures, si elles sont délicates à appliquer lorsque la taille des réseaux devient importante comme c'est le cas dans les PCRDT, apportent de nouveaux outils pour concevoir différemment les indicateurs de centralité.

Dans la sous-section suivante, nous proposons non pas un nouvel indicateur mais plutôt une nouvelle manière d'utiliser les indicateurs déjà présentés précédemment.

1.4.2.3. Une nouvelle normalisation comme indicateur complémentaire

Nous avons présenté pour chaque indicateur de centralité une mesure normalisée de manière à pouvoir comparer les mesures effectuées sur des réseaux de tailles différentes. Par exemple, si un agent participe à deux domaines technologiques au sein du même programme cadre et que l'on construit pour chacun d'eux un réseau, il peut être intéressant de comparer la place occupée par cet agent dans chaque réseau. La difficulté réside alors dans le fait que ces réseaux seront probablement de tailles différentes et ne contiendront pas les mêmes agents. Il est impossible de comparer les mesures des agents dans les deux réseaux puisque ces derniers sont de tailles différentes, la seule solution sans utiliser de normalisation est de recourir à une comparaison de type ordinal.

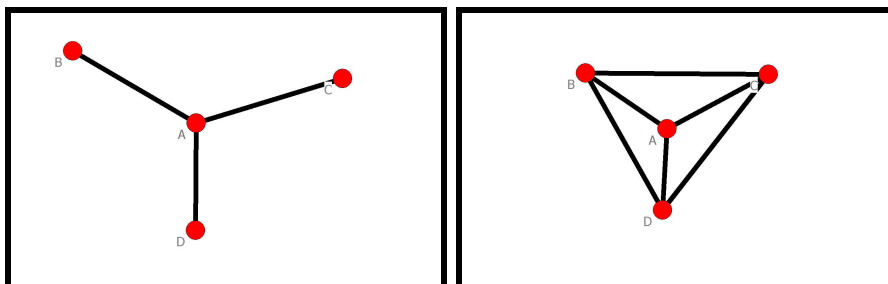
La normalisation la plus courante consiste à rapporter la mesure de centralité obtenue pour un agent dans un réseau particulier au maximum que cet indicateur peut atteindre dans ce réseau (ce maximum étant une expression de la taille du réseau). Cette uniformisation de l'indicateur présente un défaut important. Lorsque l'on compare la centralité d'un même agent dans deux réseaux différents, il ne faut pas seulement prendre en compte la mesure de l'agent relativement à ce qu'il aurait pu potentiellement faire mais aussi la performance de l'agent relativement à ce que font les autres agents dans ce même réseau. Si l'on s'en tient à la normalisation proposée, on peut arriver à conclure qu'un agent occupe des positions équivalentes au sein de deux réseaux alors qu'en réalité ce n'est pas le cas.

Il existe trois possibilités pour remédier à ce problème :

- Utiliser un indicateur complémentaire ;
- Considérer la mesure de centralité du graphe associée à l'indicateur de centralité étudié pour prendre en compte la forme de chaque réseau ;
- Compléter la normalisation proposée en introduisant une nouvelle normalisation qui compare la centralité de l'agent avec celle des autres agents du réseau.

L'objectif n'est pas de remplacer la normalisation retenue jusqu'à présent mais bien de compléter celle-ci en apportant un surplus d'information. C'est ce que nous nous proposons de faire dans cette section.

Pour illustrer nos propos, prenons l'exemple suivant. Soit les deux graphes binaires ci-dessous :



Graphes 13 et 14 : Graphes permettant d'illustrer la nouvelle normalisation.

Sur ces deux graphes, on observe que l'agent A possède dans les deux cas trois liens. Son degré est de 3 dans les deux cas.

Les deux graphes sont de tailles identiques puisqu'ils comptent chacun quatre agents. Donc, le degré normalisé de l'agent A est de 1 ou 100%. Si l'on s'en tient à cet indicateur, l'agent A occupe une position aussi centrale dans le graphe de gauche que dans celui de droite. Or, il apparaît clairement que la position de A est plus centrale sur celui de gauche puisqu'il est au centre d'un graphe en forme d'étoile, alors que, dans celui de droite, il possède autant de liens que tous les autres agents puisqu'il s'agit d'un graphe complet.

Afin de différencier la position de l'agent A entre les deux graphes, il existe deux possibilités à partir des indicateurs présentés précédemment.

Premièrement, on peut utiliser un autre indicateur de centralité. Par exemple, on pourrait utiliser le betweenness. Dans le graphe de gauche, le betweenness de l'agent A est de 1 alors que dans le second il est de 0. L'information complémentaire apportée par l'autre indicateur permet donc de discriminer la position de l'agent A entre les deux graphes. Pour cela, nous avons utilisé un second indicateur de centralité, certes complémentaire, mais qui ne repose pas sur la même définition de la centralité que le degré. Nous avons pu mettre en avant une différence mais en utilisant plusieurs indicateurs aux significations différentes. Il aurait été préférable de n'utiliser que le degré dans cet exemple.

La seconde solution consiste à utiliser le degré du réseau qui est la mesure de degré du graphe dans son ensemble. Dans le cas du graphe de gauche, cette mesure est de 100% alors que dans le graphe de droite elle est de 0%. En recoupant les informations obtenues grâce à l'indicateur de degré au niveau individuel et au niveau du graphe, nous pouvons différencier la position de l'agent A entre les deux graphes.

La difficulté réside dans le fait que le degré de centralité du graphe est une mesure qui peut s'avérer difficile à utiliser, là notamment lorsque les graphes sont de taille importante. Dans notre exemple construit de manière *ad hoc*, cela fonctionne bien. Dans la réalité, notamment dans les cas des réseaux de collaborations des PCRDT, son utilisation en complément des mesures normalisées peut s'avérer délicate surtout si les mesures pour les deux graphes que l'on cherche à comparer sont proches.

Afin de remédier à cette difficulté, nous proposons une normalisation complémentaire de la première qui utilise seulement un indicateur de centralité. Pour prendre en compte la position de tous les agents du réseau dans le but d'évaluer la performance d'un agent, nous proposons la normalisation complémentaire suivante. Soit C_i une mesure de centralité, C'_i sa normalisation et C''_i la normalisation que nous proposons.

$$C''_i = \frac{C_i}{\bar{C}_i}$$

avec \bar{C}_i la moyenne de l'indicateur de centralité retenu sur l'ensemble des agents du réseau. Si l'indicateur est supérieur à 1, c'est que l'agent est plus central que la moyenne des agents du réseau et, inversement, s'il est inférieur à 1. Si l'indicateur vaut 1, c'est que l'agent possède une mesure de centralité égale à la moyenne.

Si l'on utilise cette normalisation dans les deux graphes précédents en prenant le degré comme mesure de centralité, l'indicateur C''_A est égal à 1,5 pour le graphe de gauche et 1 pour le graphe de droite. L'agent A est donc plus central relativement aux autres agents que dans celui de droite. Il faut noter que cette normalisation est complémentaire à la précédente et ne se substitue pas à celle-ci. Il est important de conserver les deux mesures car celle proposée ci-dessus ne prend pas en compte les différences de tailles éventuelles des graphes. Elle permet d'apporter un surplus d'information.

Dans cette section, nous avons présenté de nombreux indicateurs permettant de caractériser la place qu'occupe un agent au sein du réseau. Ces indicateurs reposent sur des définitions et des intuitions différentes. Un agent peut donc être central selon un indicateur mais ne pas l'être avec un autre. Tous ces indicateurs ne sont pas forcément à utiliser simultanément. Leur utilisation dépend de ce que l'on cherche à démontrer.

Les indicateurs présentés ont pour point commun de posséder une définition de la centralité qui convient aux PCRDT. Il existe d'autres indicateurs comme le power index (Bonacich, 1987) qui peuvent être utilisés sur le réseau des PCRDT mais son apport serait plus limité pour ce cas spécifique. Le po-

wer index met en avant le fait qu'un agent qui a de nombreux liens peut être influencé négativement. Il introduit donc l'hypothèse qu'il ne faut pas forcément avoir de nombreux liens pour être central. De plus, le fait d'avoir de nombreux liens peut aussi poser des problèmes d'appropriabilité des connaissances et des informations qui circulent au sein des réseaux. Dans le cas des PCRDT, où on suppose qu'il est primordial d'être lié à de nombreuses personnes pour pouvoir acquérir de nouvelles informations et connaissances, il est difficile d'utiliser cette hypothèse d'autant plus que l'UE favorise elle-même les nombreux partenariats en finançant des projets d'envergure contenant de nombreux agents.

Dans les réseaux des PCRDT, nous supposons que circulent des connaissances, des informations et des externalités sans pouvoir réellement mesurer ces flux. Les projets n'étant pas financés entièrement, les agents ont donc un intérêt autre que le financement pour participer à ces projets. Le seul autre intérêt est de pouvoir collaborer avec d'autres agents. De fait, il est pour eux intéressant de créer un maximum de collaborations. Les indicateurs présentés ci-dessus apparaissent donc comme pertinent dans ce cadre là.

1.5 La méthode cœur/périphérie

Dans les paragraphes précédents, nous avons présenté un certain nombre d'indicateurs de centralité visant à identifier les agents centraux d'un réseau. Une seconde méthode pour déterminer les agents centraux consiste à créer une partition des agents telle que les agents centraux forment un sous-ensemble d'agents qui entretiennent des relations privilégiées entre eux.

L'objectif de cette démarche est de déterminer un sous-ensemble d'agents qui collaborent entre eux de manière étroite. Le taux de collaboration entre ces agents devra donc être supérieur à la moyenne. Dans le cadre d'un réseau de collaborations, il est intéressant d'identifier un sous-ensemble d'agents qui, par leurs relations, sont incontournables au sein du réseau. On peut supposer que ces agents, ensemble, forment alors la "plaque tournante" du réseau et que la plupart des informations et connaissances passent par eux.

L'approche cœur/périphérie est une méthode qui partitionne l'ensemble des agents en deux sous-ensembles. L'objectif est de se rapprocher d'une matrice dite "idéale". Dans un cadre idéal, les agents du cœur possèderaient des liens entre eux et des liens avec les agents de la périphérie. Ces derniers au contraire ne possèderaient que des liens avec les agents du cœur et aucun lien entre eux.

Pour présenter cette méthode, nous nous plaçons dans un premier temps dans le cadre le plus simple à savoir un réseau binaire. La matrice d'adjacence de ce réseau est divisée en 4 parties. La partie supérieure gauche décrit les relations entre les agents du cœur. Dans l'idéal, les agents du cœur doivent tous être liés les uns avec les autres d'où une densité de 100%. La partie inférieure droite représente les relations entre les agents de la périphérie qui, dans l'idéal, doivent être inexistantes d'où une densité de 0%. Les deux autres parties de la matrice renvoient aux relations entre les agents du cœur et les agents de la périphérie. Il est possible de faire plusieurs hypothèses sur ces relations. Selon ces hypothèses la matrice idéale est donc différente. La logique veut que la densité des relations cœur/périphérie soit supérieure à celle du bloc périphérie mais inférieure à celle du cœur. La taille des sous-ensembles n'est pas fixée à l'avance. Dans le cas le plus général, les relations entre les agents du cœur et de la périphérie importent peu. La matrice idéale a alors la forme suivante.

100%	-
-	0%

Afin de déterminer quels sont les agents du cœur et quels sont ceux de la périphérie, nous réalisons un important nombre d'itérations et à chaque itération la partition entre les agents du cœur et ceux de la périphérie est modifiée. Pour chaque partition, on mesure la performance réalisée à l'aide d'une mesure de "fitness" étant données les hypothèses initiales. Plus cette mesure est élevée, plus la partition se rapproche de celle de la matrice idéale. Au final, la partition pour laquelle la performance est la meilleure est retenue.

Cette approche demande de réaliser un grand nombre d'itérations. A chaque itération, l'algorithme utilisé par le logiciel Ucinet modifie la partition. Cet algorithme est un algorithme génétique. Ce dernier est utilisé en informatique et en mathématiques afin de trouver des solutions exactes ou approximatives à des problèmes d'optimisation. Cet algorithme nécessite une population de départ et une mesure de "fitness" qui permet d'évaluer la performance du résultat obtenu. L'algorithme est ici utilisé pour partitionner la population en deux sous-ensembles à chaque itération et évaluer la pertinence de la partition. D'une itération à l'autre, l'algorithme réalise des permutations pour modifier la partition. L'algorithme s'arrête soit parce qu'il a atteint un maximum pour la mesure de fitness soit parce qu'il a atteint un nombre maximal d'itérations fixées au départ. Dans ce dernier cas, au cours de l'ensemble des itérations une solution a pu ou non être atteinte. Nous modifions la matrice donnée en entrée de manière à éviter de retenir une partition qui ne serait pas optimale mais qui serait seulement la meilleure

sur l'ensemble des itérations réalisées. A noter qu'il est aussi possible que l'algorithme s'arrête sur un maximum local (Boyd, Fitzgerald et Beck, 2006) d'où l'intérêt de reproduire l'algorithme en permutant la matrice donnée en entrée (Borgatti, Everett et Freeman, 2002). Dans le cas de réseau de faible taille (comme l'exemple 1), le nombre d'itérations maximal n'est pas très élevé pour couvrir le champ des possibles. Par contre, pour les PCRDT, il est impossible de réaliser l'ensemble des permutations possibles en testant tous les sous-ensembles possibles d'agents qui est de 2^N .

Il existe plusieurs mesures de "fitness" possibles. Une première mesure est obtenue en calculant la corrélation entre la matrice soumise en entrée et la matrice idéale. En complément des informations données par la mesure de performance, il faut observer les densités des blocs cœur et périphérie. L'objectif est de trouver la partition qui maximise la densité du cœur et minimise celle de la périphérie. De plus, il faut observer aussi la répartition des agents entre le cœur et la périphérie. Par exemple, si la mesure de performance est élevée, et si les densités du cœur et de la périphérie sont respectivement proches de 100% et de 0% mais qu'un pourcentage élevé des agents sont dans le cœur (par exemple plus de 50%), il faut alors se poser la question de la pertinence de la méthode cœur/périphérie pour le réseau.

Afin d'illustrer la méthode cœur/périphérie, nous reprenons l'exemple 1. L'algorithme partitionne l'ensemble des agents de la manière suivante :

- Le cœur est composé des agents 1, 2 et 3 ;
- La périphérie est composée des agents 4, 5, 6 et 7.

Cette partition renvoie à la matrice suivante :

Agents	Ag. 1	Ag. 2	Ag. 3	Ag. 4	Ag. 5	Ag. 6	Ag. 7
Agent 1	3	1	1	1	1	0	1
Agent 2	1	1	1	0	0	0	0
Agent 3	1	1	2	0	0	0	1
Agent 4	1	0	0	1	1	0	0
Agent 5	1	0	0	1	2	1	0
Agent 6	0	0	0	0	1	1	0
Agent 7	1	0	1	0	0	0	1

$$Ag.^* = Agent$$

Cette matrice est donc comparée à une matrice idéale qui prend la forme suivante :

Agents	Ag. 1	Ag. 2	Ag. 3	Ag. 4	Ag. 5	Ag. 6	Ag. 7
Agent 1	-	1	1				
Agent 2	1	-	1				
Agent 3	1	1	-				
Agent 4				-	0	0	0
Agent 5				0	-	0	0
Agent 6				0	0	-	0
Agent 7				0	0	0	-

$$Ag.^* = Agent$$

Dans le calcul de la mesure de fitness (on utilise ici le coefficient de corrélation), les éléments de la diagonale ne sont pas pris en compte ni les relations entre les agents de la périphérie et les agents du cœur. Cette partition est celle qui obtient le meilleur fitness qui est de 0,632. L'algorithme a réalisé le nombre maximal de permutations étant donné le faible nombre d'agents. Dans le cas des PCRDT qui possèdent un nombre plus conséquent d'agents, il sera bien sûr difficile de tester l'ensemble des possibilités.

La matrice de densité obtenue pour l'exemple 1 est la suivante :

	Cœur	Périphérie
Cœur	1	0,333
Périphérie	0,333	0,333

Le fitness semble assez élevé. La densité est parfaite pour les agents du cœur puisque tous les agents du cœur sont liés les uns aux autres. La densité est faible pour les agents de la périphérie. Tous ces éléments convergent donc vers le fait que l'analyse cœur/périphérie a bien permis d'identifier le cœur du réseau. Toutefois, si l'on observe de plus près le réseau de l'exemple 1, on remarque que les agents 2 et 7 occupent une position strictement équivalente. Quel que soit l'indicateur de centralité utilisé, ils sont aussi centraux l'un que l'autre. Leur suppression du réseau engendre les mêmes modifications des propriétés générales du réseau. Pour autant, seul l'agent 2 appartient au cœur du réseau. Cet élément nous amène donc à conclure qu'il faut rejeter la partition de l'ensemble des agents obtenus même si c'est celle qui produit le meilleur résultat.

Le coefficient de corrélation est une des mesures qui peut être utilisée comme fitness mais ce n'est pas la seule. On peut aussi utiliser la densité. Si l'on utilise cette mesure sur l'exemple 1, seul l'agent 1 appartient au cœur.

L'objectif de la méthode cœur/périphérie est de mettre en avant un sous-ensemble d'agents dont les relations entre eux sont denses. Lorsque la partition obtenue ne renvoie qu'un seul agent dans le cœur, il est difficile de considérer que la partition est satisfaisante même si en observant le réseau de l'exemple 1, l'agent 1 apparaît clairement comme un agent central.

Une autre mesure de fitness peut alors être utilisée. Au lieu de s'intéresser aux relations du cœur, cette mesure prend en compte les relations des agents de la périphérie. Le fitness de la partition sera d'autant plus élevé que le nombre de liens des agents de la périphérie sera faible. Dans cette mesure, on considère seulement les liens entre les agents de la périphérie et les liens entre les agents de la périphérie et les agents du cœur. Les liens entre les agents du cœur ne sont pas pris en compte. Cette mesure appliquée à l'exemple 1 donne le résultat suivant :

- Le cœur est composé des agents 1, 2, 3 et 7 ;
- La périphérie est composée des agents 4, 5 et 6.

Cette partition renvoie à la matrice suivante :

Agents	Ag. 1	Ag. 2	Ag. 3	Ag. 7	Ag. 4	Ag. 5	Ag. 6
Agent 1	3	1	1	1	1	1	0
Agent 2	1	1	1	0	0	0	0
Agent 3	1	1	2	1	0	0	0
Agent 7	1	0	1	1	0	0	0
Agent 4	1	0	0	0	1	1	0
Agent 5	1	0	0	0	1	2	1
Agent 6	0	0	0	0	0	1	1

$$Ag.^* = Agent$$

Cette matrice est donc comparée à une matrice idéale qui prend la forme suivante :

Agents	Ag. 1	Ag. 2	Ag. 3	Ag. 7	Ag. 4	Ag. 5	Ag. 6
Agent 1	-			0	0	0	0
Agent 2		-		0	0	0	0
Agent 3			-	0	0	0	0
Agent 7				-	0	0	0
Agent 4	0	0	0	0	-	0	0
Agent 5	0	0	0	0	0	-	0
Agent 6	0	0	0	0	0	0	-

$$Ag.^* = Agent$$

La mesure de fitness est de 0,733 puisque sur les 15 liens possibles (3 entre agents de la périphérie et 12 entre un agent de la périphérie et un agent du cœur), 11 ne sont pas réalisés. La matrice de densité est :

	Cœur	Périphérie
Cœur	0,833	0,167
Périphérie	0,167	0,667

La partition obtenue semble plus cohérente dans la mesure où l'agent 2 et l'agent 7 qui occupent des places symétriques dans le réseau sont tous deux dans le même sous-ensemble. De plus, on vérifie bien que la densité du cœur est supérieure à celle de la périphérie.

Il existe d'autres mesures de fitness qui permettent d'obtenir une partition de l'ensemble des agents. Toutefois, dans le cadre de la thèse, nous ne présenterons que ces trois là pour les "one-mode networks" binaires. Quand le graphe est dense comme celui de l'exemple 1, il est préférable d'utiliser des informations supplémentaires afin de réaliser une partition plus juste. L'intensité des liens est une informations supplémentaire qui peut permettre d'obtenir une meilleure partition. De cette manière, on cherche à déterminer quels sont les agents qui du fait de leurs relations entre eux et de l'intensité de ces relations font parties du cœur. La méthode utilisée est identique à celle présentée pour les réseaux binaires au même titre que les mesures de fitness disponibles.

Il existe de très nombreux indicateurs dans l'analyse des réseaux sociaux et l'objectif de ce chapitre était de présenter ceux qui sont utiles pour notre analyse. Pour chaque indicateur, nous avons montré que l'on pouvait les employer aussi bien dans le cas des réseaux valués que dans celui des réseaux binaires (excepté pour le closeness et le betweenness). Dans le chapitre suivant, nous souhaitons construire pour chaque domaine technologique son réseau binaire et son réseau valué car nous avons montré que les indicateurs avaient des significations différentes selon le type de réseau. Étudier les deux formes de réseau paraît pertinent dans la mesure où les résultats obtenus ne sont pas redondants mais complémentaires et nous permettront donc de mieux évaluer les PCRDT.

L'analyse des réseaux sociaux apporte des informations sur la structure des réseaux de collaborations formés par les PCRDT. On peut ainsi mettre en avant les propriétés de ces réseaux et vérifier si celles-ci sont favorables à l'innovation. L'étude des propriétés générales permettra notamment de faire ressortir le maillage constitué par les projets financés. De plus, puisque nous retenons plusieurs domaines technologiques, nous pourrons vérifier si les PCRDT permettent de mettre en relation des champs différents en construisant un seul et unique réseau et, si c'est le cas, observer l'efficience du réseau

agrégé ainsi obtenu. Au sein de ces réseaux, il sera possible grâce aux indicateurs de centralité de repérer les agents importants de chaque domaine technologique par la place qu'ils occupent. En outre, l'analyse réseau permet d'observer le comportement d'un agent qui prend part aux deux domaines technologiques et le rôle qu'il joue dans l'échange entre les deux réseaux. Au total, l'analyse réseau contribue à confronter les caractéristiques des réseaux observés aux objectifs initiaux des PCRDT.

Chapitre 2

Les réseaux de collaborations des PCRDT

Les PCRDT ne sont ainsi pas seulement conçus comme un réseau de collaborations entre entreprises mais sont aussi destinés à promouvoir les collaborations entre les centres de recherche, les agents de l'enseignement supérieur d'un côté et les entreprises de l'autre. Pour fonder ce choix, on peut notamment citer l'article de Jaffe (1989) qui montre que collaborer avec des universités est bénéfique aux entreprises en termes de dépôts de brevets. Une autre catégorie d'agents que l'UE veut impliquer dans les PCRDT est les Petites et Moyennes Entreprises (PME). Malgré leurs qualités, ces entreprises ont souvent des difficultés à développer des activités de R&D de grande ampleur ou à participer à des projets d'envergure en R&D et ainsi à accéder aux résultats de la R&D. A. Tholoniati (2008), en s'appuyant sur les chiffres du Conseil Économique et Social Régional Rhône-Alpes, écrit : *Le Conseil Économique et Social Régional Rhône-Alpes (2006) fait remarquer que les grandes entreprises en France représentent 3% des entreprises qui font de la R&D alors qu'elles concentrent 69% des dépenses de R&D. Les PME ont un rôle non négligeable dans le processus d'innovation. Du fait de leur petite taille, elles sont souvent plus réactives et plus flexibles face à des perturbations. Il faut les préserver tout en les incitant à collaborer avec les groupes bien qu'elles manquent de stratégie, et avec les laboratoires bien qu'elles aient, originellement, de très faibles liens avec eux. En parallèle, les grandes entreprises et les groupes ont des moyens financiers, humains et matériels beaucoup plus importants. Certes, cette remarque concerne la France mais illustre parfaitement les capacités que l'on prête aux PME. L'UE est consciente des enjeux d'intégrer ces entreprises aux PCRDT : "About 25 millions SMEs account for approximately two-thirds of Europe's employment and GDP. Therefore, it comes as no surprise that SMEs are a key component*

of the research and innovation system and that they play an important role in transforming knowledge into new products, processes and services. Due to their ability to adapt quickly to market changes, SMEs are often better positioned to exploit new and emerging research opportunities that address ongoing social, environmental and economic challenges”.

De plus, si cet EER doit être un espace de collaborations permettant de rapprocher les agents les uns des autres en termes de distance dans les réseaux sociaux, comme ceux formés par les PCRDT, il a aussi pour but de faire collaborer des agents de pays différents qui n’auraient pas forcément d’autres opportunités en dehors des PCRDT. C’est le cas en particulier des agents localisés dans les nouveaux États membres de l’UE. En effet, le 6^{ème} PCRDT se déroule de 2002 à 2006 soit en plein processus d’adhésion de ces nouveaux États membres. L’UE accorde donc une importance particulière à l’intégration de ces pays membres afin d’éviter de trop grands écarts au sein des pays de l’UE. Si l’objectif premier des programmes cadres est de rattraper le retard en matière de R&D entre l’Europe d’un côté et le Japon et les États-Unis de l’autre, l’Union Européenne souhaite que ce rattrapage se fasse de manière harmonieuse au sein de l’UE comme l’avait indiqué le Centre de Recherche Scientifique et Technique (CREST) pour qui un développement scientifique et technique équilibré au sein de la communauté est important.

Il est possible grâce à l’analyse réseau d’étudier les propriétés des réseaux formés par les PCRDT et apprécier si ces propriétés sont favorables à l’innovation. Nous nous intéressons ainsi aux différentes propriétés présentées dans les chapitres précédents. Plus particulièrement, nous cherchons à déterminer si les réseaux des PCRDT sont connexes et si les agents sont proches les uns des autres au sein des réseaux. Au travers de l’étude de ces propriétés générales, nous montrons si ces propriétés correspondent à celles de l’une des formes de réseau connues pour être un réseau favorable à l’innovation : les “small world networks”. Dans ce but, nous étudierons la résistance des réseaux des PCRDT à la suppression des “hubs”, c’est-à-dire des agents qui possèdent un nombre important de liens. Ensuite, nous vérifierons si les réseaux formés par les PCRDT impliquent de nombreux agents et si ces agents collaborent les uns avec les autres. De plus, il est nécessaire d’identifier les agents centraux au sein des réseaux notamment ceux qui occupent une position de relais (au sein d’un domaine technologique, entre différents domaines technologiques ou entre les PCRDT) qui favorisent la diffusion des connaissances. L’analyse des réseaux permet également d’observer le rôle de certains agents au sein des réseaux de collaborations comme les PME ou les agents issus des nouveaux États membres.

Cependant, l’étude des réseaux de collaborations individuelles ne suffit

pas à étudier les PCRDT. Dans la section 2.3, nous construisons des réseaux de relations entre régions européennes afin de rendre compte de la dimension géographique des relations établies au sein de l'EER. Si les réseaux de collaborations individuelles et les réseaux de relations entre régions sont construits à partir des mêmes données, les informations apportées par chaque réseau sont différentes. En plus d'identifier les agents centraux, il est important de déterminer les régions motrices en matière d'innovation et les axes de collaborations privilégiés en Europe. L'EER se veut aussi un espace de collaborations entre des agents de nationalité différente. Un des objectifs du 6^{ème} PCRDT est d'intégrer des agents issus des nouveaux États membres et d'intégrer ces pays et donc leurs régions. Les réseaux de régions permettront davantage de vérifier cet objectif que les réseaux de pays qui sont trop agrégés et risqueraient de masquer une certaine hétérogénéité au sein même des pays. Les indicateurs utilisés pour étudier les réseaux de collaborations entre agents seront adaptés pour étudier les réseaux de régions. Enfin, si l'on étudiait l'ensemble des thématiques, nous pourrions observer s'il existe des spécificités régionales.

Dans la section 2.1, nous chercherons à déterminer si les réseaux de collaborations des deux domaines technologiques présentent des propriétés qui favorisent l'innovation. De plus, nous essaierons d'identifier quels sont les agents centraux des deux réseaux et quel est l'impact de ces agents sur le réseau. Dans la section 2.2, nous chercherons à montrer que les réseaux des PCRDT sont des réseaux qui se structurent et que d'un programme cadre à l'autre on tend vers un espace de recherche intégré et de plus en plus dense. Pour cela, nous étudierons un sous-domaine qui est celui des Nanotechnologies et nous établirons des comparaisons entre le réseau formé dans le 5^{ème} PCRDT et celui formé dans le 6^{ème} PCRDT. La section 2.3 permettra d'étudier les réseaux des PCRDT au niveau plus agrégé des régions.

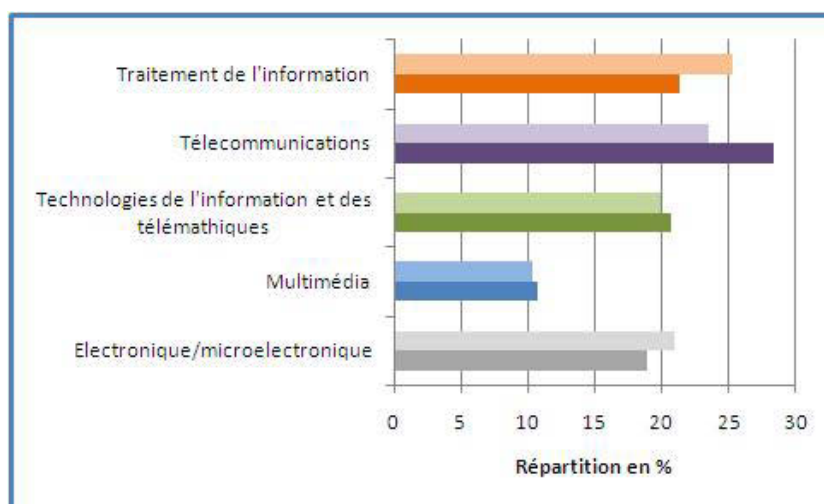
2.1 Des réseaux de collaborations individuelles aux propriétés favorables

Dans cette section, il aurait été intéressant de pouvoir étudier l'ensemble des PCRDT ou au moins tout le 6^{ème} PCRDT. La difficulté réside alors dans l'identification des agents (la recherche de doublons) et l'harmonisation des données comme décrits précédemment dans la thèse (cf. Annexes A.3). Pour contourner cette difficulté et améliorer la qualité des données à notre disposition, nous étudions seulement les domaines des Télécommunications

et de l'Electronique-Microélectronique qui sont deux domaines appartenant à la thématique IST du 6^{ème} PCRDT. Étudier deux domaines technologiques en parallèle, qui plus est des domaines proches, permettra d'avoir une vue plus globale des PCRDT et de limiter les risques d'avoir des conclusions spécifiques à un domaine. Ce choix permettra notamment de voir s'il existe des connexions entre ces deux domaines, connexions réalisées par les agents qui participent aux deux domaines. De la même manière, il sera intéressant de vérifier si, d'un domaine à l'autre, des collaborations se répètent.

2.1.1 Présentation des deux domaines technologiques

La thématique IST est celle qui reçoit le plus de financement parmi les 7 thématiques. Le montant des financements est de 3984 millions d'euros sur les 12438 millions d'euros attribués aux 7 priorités thématiques¹. La thématique IST est décomposée de 5 domaines technologiques. La répartition des projets et des participations entre ces domaines technologiques est donnée dans le graphique ci-dessous.



Graphique 1 : Répartition des projets (en clair) et des participations (en foncé) du domaine IST entre les 5 domaines technologiques.

Sur ce graphique, nous observons que c'est le domaine des Télécommunications qui compte le plus de participations. Il est donc intéressant pour la suite de conserver ce domaine technologique. Le second domaine choisi est de celui de l'Electronique-Microélectronique. Ce choix peut surprendre dans la

¹Le budget total du 6^{ème} PCRDT est de 17883 millions d'euros (cf. annexe A.1. pour plus d'informations sur les PCRDT).

mesure où les deux domaines sont de taille différente. Toutefois, ce domaine présente l'avantage d'être structuré de manière relativement identique dans le sens où les proportions de chaque type d'agent et de chaque type d'instrument sont relativement similaires entre les deux domaines technologiques.

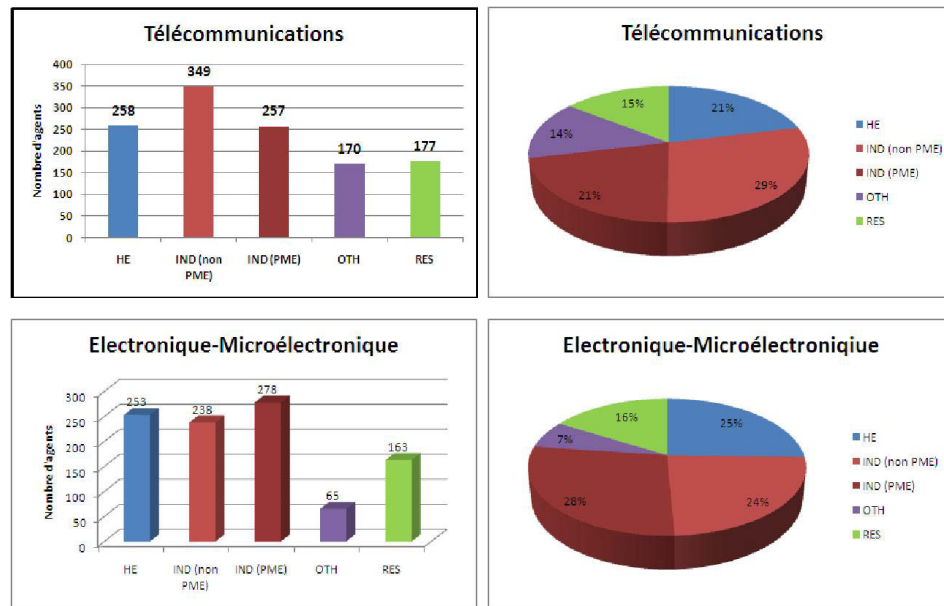
Taille des domaines. Le domaine Télécommunications compte 20% d'agents en plus et 50 % de participations supplémentaires. Ceci s'explique par la différence des financements attribués dans chaque domaine. Les principaux chiffres résumant les deux domaines technologiques sont donnés dans le tableau ci-dessous.

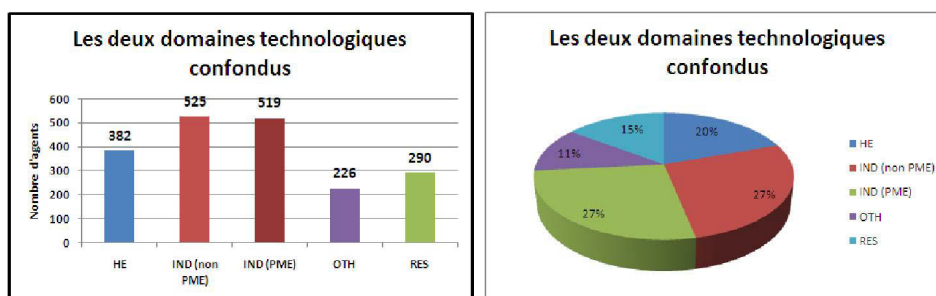
Indicateurs	Tel*	EM**	DDC***
Nombre d'agents	1211	997	1942
Nombre de projets	187	167	354
Nombre de participations	2979	1969	4948

(* *Tel* = Télécommunications ; ** *EM* = Electronique-Microélectronique ; *** *DDC* = Les deux domaines confondus.)

Tableau 8 : Répartition des agents dans chaque domaine.

Le graphique ci-dessous met en avant la répartition des agents dans chaque domaine technologique en fonction de leur type. Pour rappel, les types des agents sont : *IND* Industries, *HE* Agents de l'enseignement supérieur, *RES* Centres de recherche et *OTH* Autres.





Graphiques 2 à 7 : Répartition des agents en fonction de leur type et du domaine technologique.

L'histogramme donne le nombre d'agents par type en valeur absolue tandis que le graphique circulaire donne la répartition afin de pouvoir comparer les deux domaines technologiques.

Type 1	Intersection
HE	129
IND (non PME)	62
IND (PME)	16
OTH	9
RES	50
Total	266

Tableau 9 : Nombre d'agents communs aux deux domaines technologiques.

La répartition des agents en fonction de leur type est relativement proche. Le domaine Electronique-Microélectronique présente un pourcentage de PME plus important que le domaine Télécommunications (28% contre 21 %). Toutefois, dans les deux cas, les entreprises (PME ou non PME) représentent plus de 50% des agents. Une autre différence mineure à noter est la place prépondérante prise par les agents de la catégorie autres dans les Télécommunications.

Par ailleurs, il est intéressant d'identifier les agents qui participent aux deux domaines technologiques simultanément. En fait, ces agents sont au nombre de 266 et représentent 14% du total des agents tous domaines technologiques confondus. Si on prend les domaines technologiques séparément, les agents communs représentent 22% des agents du domaine Télécommunications et 27% de ceux du domaine Electronique-Microélectronique. Près de la moitié de ces agents sont issus de l'enseignement supérieur, catégorie

composée en grande partie d'universités. Les universités sont par définition pluri-disciplinaires ce qui explique en grande partie ce chiffre. De plus, même si l'on prend seulement en compte les universités scientifiques par exemple, il est courant d'observer l'existence de plusieurs laboratoires de recherche au sein de l'université avec des spécialités différentes, laboratoires qui peuvent aussi avoir plusieurs axes de recherche. Combiner au fait que les domaines technologiques sont proches, ce chiffre apparaît logique. Ensuite, on retrouve les entreprises non PME notamment les groupes qui ont des activités diversifiées mais aussi les grandes entreprises qui développent des produits à partir de plusieurs technologies et qui ont d'autant plus intérêt à participer aux deux domaines technologiques que ceux-ci sont proches. Le nombre de PME qui participent aux deux domaines technologiques est faible ce qui s'explique de deux manières : premièrement, la faible capacité de ces agents à participer à plusieurs projets simultanément (cf. sous-section 2.1.3) et, deuxièmement, le fait que les PME ont souvent une activité plus spécialisée et sont peu dispersées.

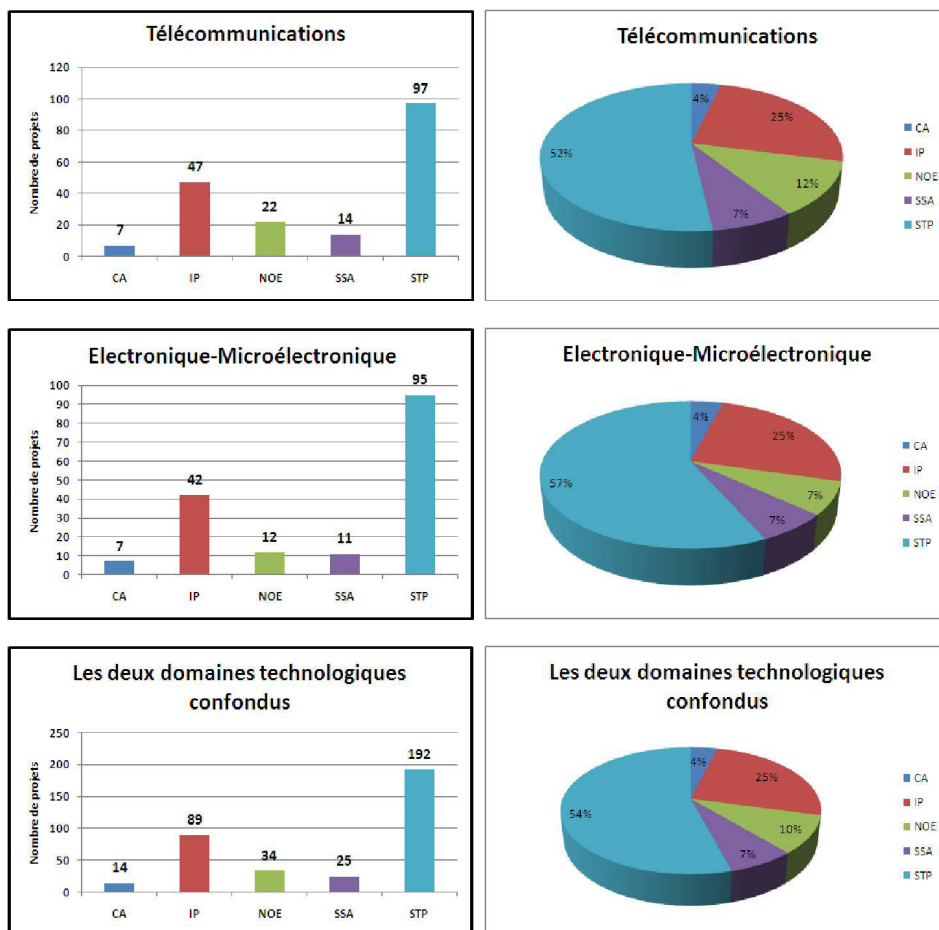
Répartition des projets au sein de chaque domaine. Le graphique ci-dessous présente la répartition des projets au sein de chaque domaine technologique². Il existe 5 instruments différents au sein des PCRDT (5 formes de projets) :

- Network of Excellence (NoE) ;
- Integrated Projects (IP) ;
- Specific Targeted REsearch Projects (STREP) ;
- Coordination Action (CA) ;
- Specific Support Action (SSA).

Les deux premières formes de projets sont des instruments qui visent à ce que les groupes d'agents ainsi formés atteignent la masse critique suffisante afin de mener à bien leur projet de recherche. Ces deux instruments correspondent à des projets d'envergure qui rassemblent un nombre important d'agents. Les projets STREP sont moins "ambitieux" que les projets NOE et IP et de taille plus réduite. Ces projets visent à améliorer la compétitivité de l'UE et cherchent à cibler les besoins de la société actuelle. Ils peuvent prendre la forme d'un projet de R&D visant soit à accroître le champ des connaissances dans un domaine soit à s'assurer de la viabilité d'une technologie qui ne peut pas être commercialisée dans l'état actuel des choses. Les projets SSA et CA possèdent des objectifs différents des précédents instruments. Ils sont de plus faible envergure et sont des points d'entrée pour les

²Cf. Annexes A.2 pour une définition plus précise des différentes formes de projet

structures de faible taille ou ont pour vocation de préparer des projets futurs via l'organisation de conférences par exemple.



Graphiques 8 à 13 : Répartition des projets en fonction des instruments et du domaine technologique.

L'histogramme donne le nombre de projets par instrument en valeur absolue tandis que le graphique circulaire donne la répartition afin de comparer les deux domaines technologiques.

On remarque que plus de 50% des projets sont des projets de type *STP* pour les deux domaines technologiques. Les projets d'envergure de type *NOE* et *IP* représentent 32% des projets financés dans le domaine Electronique-Microélectronique et 37% dans le domaine Télécommunications. Le pourcentage de projets de type *IP* est identique. Par contre, le domaine des Télécommunications compte 12% de projets *NOE* contre seulement 7% pour le

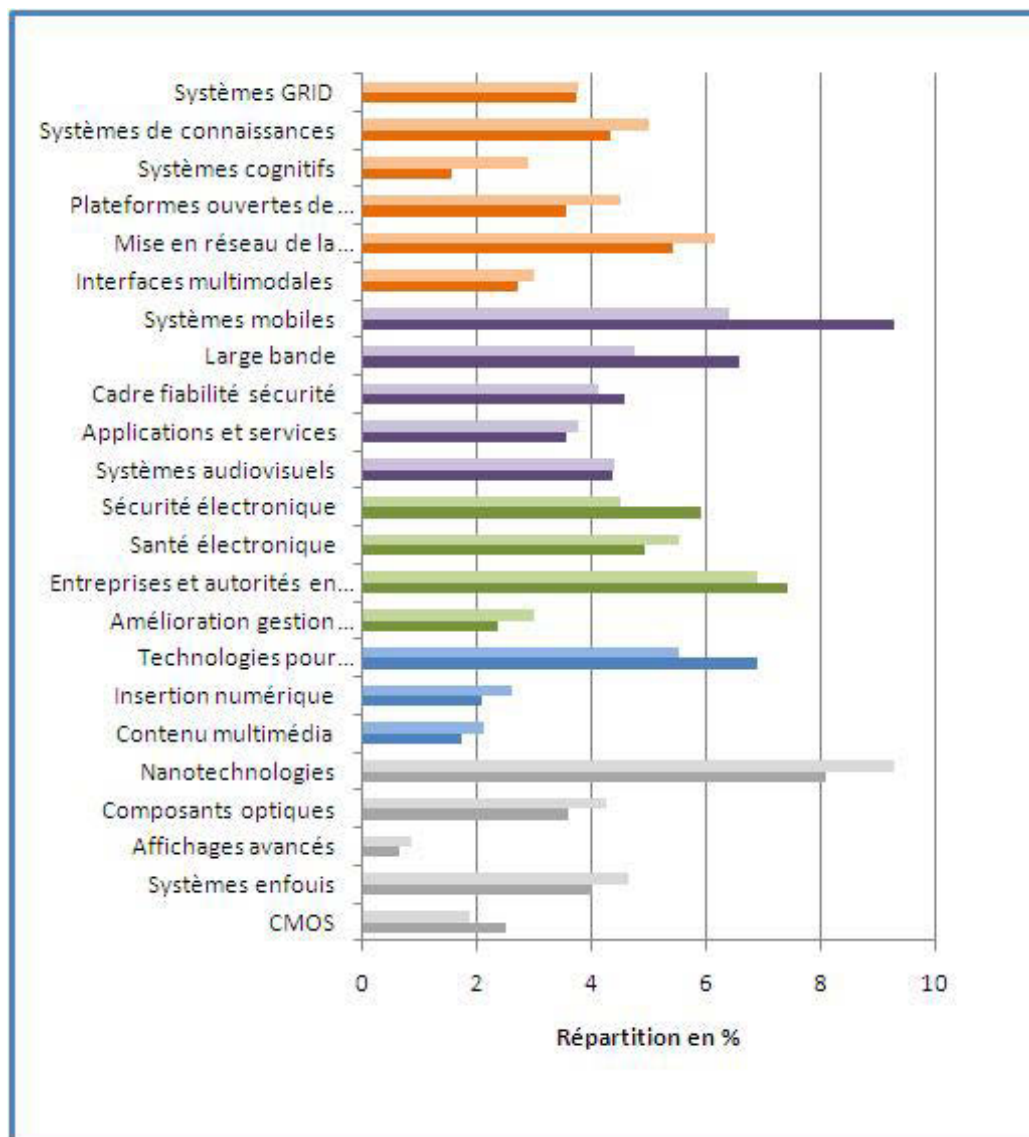
domaine Electronique-Microélectronique. Le fait que les projets *NOE* soient en moyenne les projets de plus grande taille (cf. tableau ci-après) peut expliquer le nombre d'agents plus important dans le domaine Télécommunications ainsi que le nombre de participations et de liens plus importants. De plus, sur le tableau ci-dessous, on observe que la taille des projets est plus importante dans le domaine des Télécommunications, et ce, quels que soient les instruments (sauf CA).

INST	Electronique-Microélectronique		Télécommunications	
Instrument	Nombre moyen d'agents par projet	Ecart-type	Nombre moyen d'agents par projet	Ecart-type
CA	14,43	12,29	10,29	5,68
IP	18,48	8,66	27,85	9,36
NOE	28,92	9,47	30,41	14,75
SSA	4,91	3,05	5,50	4,75
STP	7,27	1,92	8,78	2,48
Global	11,79	8,95	15,93	12,20

Tableau 11 : Nombre d'agents par projet dans les deux domaines technologiques.

Electronique-Microélectronique et Télécommunications. Enfin, le fait de retenir l'Electronique-Microélectronique comme second domaine présente aussi un autre avantage. Chaque domaine est composé en sous-domaines. Or, parmi les sous-domaines de l'Electronique-Microélectronique, on trouve les nanotechnologies, sous-domaine pour lequel nous possédons les collaborations dans le 5^{ème} PCRDT et le 6^{ème} PCRDT et qui est aussi le sous-domaine de la thématique IST du 6^{ème} PCRDT qui possède le plus grand nombre de projets comme le montre le graphique ci-dessous. En effet, nous avons vu qu'il n'existait pas de stricte correspondance dans le temps entre les différentes thématiques d'un programme à un autre, limite qui rendait délicate une étude dans le temps des PCRDT. Le fait de posséder un sous domaine pour lequel la continuité entre deux programmes cadres est possible, permettra par la suite de compléter notre analyse statique par une analyse plus dynamique et de montrer dans quelle mesure l'analyse réseau peut rendre compte de l'évolution du réseau de collaborations des PCRDT entre deux périodes (cf. section 2.2). Enfin, nous montrerons aussi dans quelle mesure l'analyse réseau peut être utile dans l'étude des déterminants des choix de collaborations et en quoi, dans ce cas précis, elle est utile dans les estimations

économétriques (cf. Partie II).



Graphique 14 : Répartition des projets (en clair) et des participations (en foncé) du domaine IST entre les sous-domaines technologiques. La couleur de chaque sous-domaine renvoie au domaine technologique dont il fait partie.

Les premières conclusions que l'on peut tirer indiquent que ces deux domaines technologiques attirent un nombre important d'agents et des agents différents. On observe notamment que les entreprises constituent 50% des

agents. De plus, parmi ces 50%, il est important de noter la place prépondérante prise par les PME. L'objectif de l'UE d'attirer des agents diversifiés qui ont une vision différente de la recherche (plus fondamentale pour les agents de l'enseignement supérieur et plus appliquée pour les entreprises) et de voir participer un grand nombre de PME semble donc atteint. Toutefois, il reste à vérifier le rôle que joue chaque type d'agent dans le réseau et notamment la place réelle prise par les PME. Dans cette section, nous chercherons à déterminer quels sont les agents centraux des réseaux de collaborations propres à chaque domaine mais auparavant il est nécessaire d'étudier les propriétés des réseaux et de vérifier si celles-ci sont favorables à l'innovation. Pour un agent, participé aux PCRDT et occupé une position centrale au sein du réseau ne sont intéressants que si le réseau est propice à l'innovation et si le réseau forme réellement, selon la volonté de l'UE, un EER. C'est l'objet de la sous-section suivante.

2.1.2 Les réseaux des PCRDT possèdent-ils des propriétés favorables à l'innovation ?

L'étude des propriétés générales des réseaux permet de donner une vue d'ensemble du réseau et de déterminer si le réseau possède des propriétés susceptibles de favoriser l'innovation. Une manière de procéder est de chercher à montrer que les réseaux étudiés possèdent des propriétés similaires à une forme de réseau particulière reconnue pour posséder de "bonnes propriétés" et, plus précisément, des propriétés qui favoriseraient l'innovation.

Dans le cas des réseaux de collaborations en innovation, les "small world networks" (ou "réseaux petit monde") sont souvent présentés comme ayant de "bonnes propriétés". Ce réseau est un type de graphe dans lequel les agents collaborent peu les uns avec les autres mais peuvent communiquer (au sein du réseau) à travers un nombre relativement faible de liens. Ces réseaux mettent en avant le fait que des agents qui ne collaborent pas ensemble sont liés indirectement par un partenaire commun. Ces réseaux présentent des propriétés particulièrement intéressantes pour des collaborations en innovation puisqu'ils mettent en avant une certaine efficacité du réseau dans la transmission des connaissances et des informations qui est bénéfique aux activités d'innovation. C'est encore plus vrai lorsque l'on suppose que les hypothèses de "decay" ou de réseau borné ont cours dans le réseau. Nous définissons plus précisément ce type de réseau dans le paragraphe suivant.

2.1.2.1. Définition d'un "small world network"

Les propriétés propres aux "small world network" ont été mises en avant par Watts et Strogatz (1998) et Watts (1999). Ils ont remarqué que de nombreux réseaux réels (par opposition aux réseaux aléatoires³) sont caractérisés par une faible distance entre agents au sein du graphe et un coefficient de cluster élevé. C'est le cas d'internet, de réseaux neuronaux, de réseaux de co-auteurs et de nombreux autres réseaux réels.

Granovetter (1973) a ainsi souligné l'importance des liens faibles (ce sont des liens qui relient deux sous-ensembles d'agents, sous-ensembles qui sans ce lien n'auraient pas de liens directs entre eux ou alors seulement un lien indirect qui induirait une distance plus élevée). Ces liens ont pour principale incidence de diminuer la distance et le diamètre au sein du réseau. Cela a pour principale conséquence d'éviter que les agents soient exclus du réseau ou soient situés à une distance élevée des autres agents. La proximité d'un sous-ensemble d'agents avec les autres sous-ensembles permet d'éviter à ces agents de fonctionner entre eux (en vase clos) et, au contraire, d'être ouverts aux autres et ainsi de pouvoir profiter de leurs idées et influences. Granovetter cite ainsi l'exemple du marché du travail et le fait que ces liens permettent d'avoir accès à des offres de travail auxquelles un agent n'aurait pas eu accès s'il était lié uniquement avec un sous-ensemble plus restreint d'agents. Dans le cadre d'un réseau de collaboration en R&D, ces liens faibles amènent de la diversité et un renouvellement d'idées. Sans étudier les liens faibles au sein des PCRD, des distances faibles et un réseau connexe permettent ainsi de pouvoir interagir plus facilement avec les autres agents et de pouvoir profiter de leurs idées, travaux et informations afin d'améliorer ses propres travaux de recherche.

Il existe plusieurs définitions des propriétés d'un "small world network". Dans cette sous-section, nous en proposerons deux : celles de Newman, Watts, Strogatz et celles de Goyal, van der Leij et Moraga-Gonzalez. Nous verrons que si ces deux définitions retiennent des indicateurs identiques pour définir les "small world networks", elles n'ont pas les mêmes points de comparaison.

La définition de Newman, Watts, Strogatz Selon Watts et Strogatz (1998), deux indicateurs sont pertinents pour établir qu'un réseau est un

³Un réseau aléatoire est généré par un processus aléatoire. Pour construire un tel réseau, il faut partir d'un ensemble d'agents N et construire les liens entre eux de manière aléatoire. Il existe différents graphes aléatoires basés sur des distributions de probabilité différentes. L'un des modèles les plus étudiés est le modèle d'Erdős-Rényi (1959, 1960).

“small world network” : c’est la distance moyenne au sein du réseau et le coefficient de cluster du réseau. De manière à pouvoir établir que le réseau étudié est un “small world network”, Watts propose de comparer ces deux indicateurs à ceux obtenus dans des réseaux aléatoires de même taille.

Pour déterminer si un réseau est un “small world network”, Watts (1999) compare le coefficient de cluster et la distance moyenne obtenus dans le réseau réel à ceux obtenus dans un réseau aléatoire de même taille. Plus le rapport est proche de 1 dans le cadre de la distance moyenne et excède 1 dans le cas du coefficient de cluster, plus le réseau présente les propriétés d’un “small world network”.

Dans notre cas, la difficulté est que notre réseau est un graphe bipartite, c’est-à-dire un réseau d’affiliation et que nos “one-mode networks” sont obtenus par projection à partir du “two-mode network”. Par construction, le coefficient de cluster d’un agent qui ne participe qu’à un projet est de 1 et de fait cela aura tendance à biaiser le coefficient de cluster du réseau. Newman, Watts et Strogatz (2001) introduisent un moyen simple de calculer le coefficient de cluster d’un réseau d’affiliation mais aussi d’un réseau aléatoire comparable à un réseau d’affiliation.

Par la suite, nous utiliserons les deux indicateurs de distance moyenne et de coefficient de clustering pour déterminer si les agents des réseaux des PCRDT sont plus proches les uns des autres que si l’appariement entre les agents se faisait de manière aléatoire.

Soient N l’ensemble des agents avec $|N| = n$ et M l’ensemble des projets avec $|M| = m$. Soient μ le nombre de projets moyen par agent et ν le nombre d’agents moyen par projet. Newman, Watts et Strogatz (2001) supposent que ces deux distributions suivent une loi de Poisson. Ils démontrent que la distance moyenne (DM_{Alea}) et le coefficient de cluster (CCl_{alea}) d’un réseau aléatoire comparable à un réseau d’affiliation (ou bipartite graph) peuvent être obtenus de la manière suivante :

$$DM_{Alea} = \frac{\ln(n)}{\ln(\mu \cdot \nu)}$$

$$CCl_{Alea} = \frac{m \nu^3}{n \nu^2(\mu^2 + \mu)} = \frac{1}{1 + \mu}$$

Ces deux indicateurs serviront alors d’étalon pour évaluer les propriétés des réseaux des PCRDT.

La définition de Goyal, van der Leij et Moraga-Gonzalez Goyal, van der Leij et Moraga-Gonzalez (2006) mettent en avant un certain nombre de

critères qui permettent de vérifier si un réseau est un “small world network”. Leur définition d’un tel réseau est différente de celle proposée par Watts (1999). Les critères retenus par Goyal, van der Leij et Moraga-Gonzalez pour déterminer si un réseau est un “small-world network” sont les suivants :

- Le nombre d’agents n est très important par rapport au nombre de liens moyen par agent (\overline{CD}) tel que $n \gg \overline{CD}$;
- Le réseau est intégré, c’est-à-dire qu’une grande composante existe et couvre une large part des agents ;
- Le coefficient de cluster CCl est tel que $CCl > \overline{CD}/n$
- La distance moyenne (DM) entre les agents de la grande composante est faible, DM est de l’ordre de $\ln(n)$ ⁴.

Au paragraphe suivant, nous examinerons si les réseaux des PCRDT satisfont ces critères.

Enfin, pour compléter les deux définitions ci-dessus et être certains que les réseaux sont des “small world networks”, nous vérifierons si les réseaux suivent une loi de puissance qui est une propriété souvent associée aux “small world networks”. Nous définissons cette loi dans le paragraphe suivant.

Loi de puissance Une loi de puissance exprime une relation entre deux variables x et y qui n’est pas linéaire. De fait, ajuster la relation entre les variables x et y par une régression linéaire serait de fait une erreur. La loi de puissance décrit la relation entre les deux variables de la manière suivante :

$$y = \alpha x^k$$

avec α une constante dite “constante de proportionnalité” et $|k|$ le degré de la loi. La valeur de k est négative. Appliqué aux réseaux sociaux, on suppose alors que la probabilité qu’un agent ait un degré x est :

$$P(x) = \alpha . x^k$$

La relation décrite ci-dessus peut se réécrire :

$$\log(y) = k.\log(x) + \log(\alpha)$$

⁴Cohen et Havlin (2003) ont montré notamment que les “small world networks” possédaient un diamètre très faible $D \sim \ln(\ln(n))$ alors que Goyal, van der Leij et Moraga-Gonzalez (2004) retiennent comme critère $D \sim \ln(n)$.

Sur un graphique utilisant des échelles logarithmiques, le graphe d'une loi de puissance est une droite. Dans le cas des "small world network", de nombreux articles (Reka et Barabasi, 2002 ; Adamic et Huberman, 2002 ; Jarvis, 2006 ; Roediger-Schluga et Barber, 2008) montrent que le coefficient de pente est compris entre -3 et -2.

Dans les réseaux sociaux, ce type de distribution apparaît dans les "small world networks", car de nombreux agents possèdent peu de liens et il existe des "hubs" en faible nombre qui concentrent une grande part des liens. On se rapproche souvent dans ce cas de la loi "80/20" ou loi de Pareto. Cette loi appliquée aux réseaux sociaux indique que 80% des agents possèdent 20% des liens et que les 20% des agents restant concentrent les 80% des liens restant. La difficulté à appliquer ce concept aux participations et aux liens au sein des réseaux des PCRDT est le fait qu'un projet renvoie à plusieurs agents et qu'un lien implique exactement deux agents. De ce fait un lien est comptabilisé deux fois, une fois pour chaque agent ce qui rend plus difficile l'apparition d'une loi 80/20.

La distribution des liens est alors représentée sur un graphique avec en abscisse un nombre de liens et en ordonnée le nombre d'agents qui possèdent ce nombre de liens. Il est possible de transformer le graphique en utilisant pour les deux axes des échelles logarithmiques et ainsi modifier l'allure du nuage de points. Si la distribution étudiée suit une loi de Poisson alors les points doivent être alignés et donner visuellement l'impression de pouvoir être reliés par une même droite, droite qui aurait pour équation $\log(y) = k \cdot \log(x) + \log(\alpha)$.

Généralement dans l'étude des réseaux, seule la distribution des liens est utilisée. Appliquée aux réseaux formés par les PCRDT, la pertinence de cet indicateur est problématique. En effet, chaque participation à un projet pour un agent implique (presque dans tous les cas) la création d'au moins deux liens puisque les projets sont composés d'au moins trois agents. Précédemment, nous avons indiqué que la taille moyenne des projets était d'environ 12 agents dans le domaine Electronique-Microélectronique et 16 agents dans le domaine Télécommunications. En conséquence, il existe peu d'agents qui possèdent un nombre de liens compris entre 1 et 10. Il est ainsi utile de compléter l'analyse de la distribution des liens par celle de la distribution des participations.

L'étude des distributions des liens et des participations nous fournit de premières indications sur l'existence de "hubs" au sein des réseaux, c'est-à-dire d'agents centraux dans le sens où ils possèdent un nombre très important

de liens (dans notre cas, nous utiliserons la mesure de degré pour identifier les “hubs”). Afin de vérifier l’existence de “hubs”, nous utilisons en complément l’indice de Gini qui permet d’étudier si les liens sont répartis de manière équitable entre les agents.

De nombreuses autres propriétés sont associées aux “small world networks”. Les “small world networks” sont caractérisés aussi par un nombre important de cliques⁵ ou de sous-graphes, ces cliques ou sous-graphes étant eux-mêmes connectés par un nombre restreint de liens. La présence de “hubs” est aussi un élément permettant de repérer l’existence d’un “small world network”. Si la distribution du degré des agents suit une loi de puissance alors cela peut être interprété comme un indice pour dire que le réseau est un “small world network”. Un “small world network” qui suit une loi de puissance est robuste à la suppression aléatoire d’un agent, c’est-à-dire que la suppression aléatoire d’un agent modifie peu les propriétés du réseau. Cela s’explique par le fait que de nombreux liens passent par les “hubs” du réseau. Par contre, si un “hub” est supprimé alors cela peut causer de fortes modifications des propriétés du réseau et, de fait, nuire à son efficacité. En comparaison, la suppression d’un agent quelconque dans un réseau aléatoire peut induire une augmentation assez importante de la distance moyenne dans la mesure où, dans les réseaux aléatoires, les agents ont un nombre de liens relativement équivalent.

Lorsque l’on observe l’ensemble des propriétés d’un “small world network”, il apparaît clairement que si les réseaux des PCRDT s’en rapprochent, alors une grande partie des objectifs affichés par la politique européenne des PCRDT sera vérifiée. Une grande composante, voire un réseau entièrement connexe, combinée à une faible distance entre les agents serait le signe d’un réseau intégré et efficace. De plus, un réseau robuste à la suppression des agents centraux est un réseau qui est moins dépendant des agents qui le composent et qui peut assurer des transmissions efficaces de flux volontaires (informations et connaissances) ou involontaires (externalités de connaissances) au sein du réseau. La présence ou l’absence de “hubs” peut être interprétée de façon diverse au regard de la politique souhaitée par l’UE. En effet, l’apparition de “hubs” peut aller à l’encontre de la politique souhaitée par l’UE. En effet, si un nombre restreint d’agents concentre une trop grande partie des liens, il est aussi possible qu’il concentre une grande partie des financements et des participations ce qui serait contraire à la volonté de l’UE d’attirer de nombreux agents. Toutefois, afin de faciliter la diffusion des connaissances

⁵Une clique est un sous-ensemble d’agents tel qu’ils ont tous liés directement les uns avec les autres.

au sein du réseau et d'assurer une meilleure efficacité de celui-ci, la présence de "hubs" est indispensable. Ces agents assurent souvent un rôle de relais qui permet de favoriser la circulation des flux au sein d'un réseau. Malerba, Vonortas, Breschi et Cassi (2006) définissent un "hub" de la manière suivante : *a node with a large number of connections or, alternatively, as a node that is highly influential by playing the role of network connector, i.e. one connecting nodes that would otherwise remain unconnected*. La définition proposée ici est une définition précise, de nombreuses études s'en tiennent à la première partie de celle-ci. On observe que certains objectifs peuvent être difficilement conciliables puisqu'en voulant favoriser l'innovation au sein des réseaux des PCRDT, l'UE doit pour cela s'appuyer sur des "hubs" qui risquent de concentrer les collaborations et donc les financements.

Dans ce cas précis, l'analyse réseau permet d'améliorer les connaissances sur les réseaux de collaborations formés dans le cadre des PCRDT tout en ayant une application directe qui est l'appréciation partielle de la concordance avec les objectifs fixés aux PCRDT. Dans le paragraphe suivant (2.1.2.2.), nous étudions les propriétés générales des réseaux. Puis, nous vérifierons si les réseaux suivent une loi de puissance (paragraphe 2.1.2.3.). Enfin, nous examinerons si les réseaux sont robustes à la suppression des "hubs" (paragraphe 2.1.2.4.).

2.1.2.2. Propriétés générales des réseaux

A partir des collaborations entre agents, nous avons construit un réseau pour chaque domaine technologique ainsi qu'un réseau agrégé de ces deux domaines. Pour chacun de ces réseaux, nous avons calculé leurs propriétés générales. Ces propriétés générales sont données dans le tableau ci-dessous.

Indicateurs	TEL^1	EM^2	DDC^3
Nombre d'agents	1211	997	1942
Nombre de projets	187	167	354
Nombre de participations	2979	1969	4948
Nombre de liens	26274	15462	40394
Valeurs totales des liens	36074	17277	53351
Nombre potentiel de liens	732655	496506	1884711
Densité (en %)	3,59	3,11	2,14
Degré moyen (réseau binaire)	43,4	31	41,6
Intensité moyenne	1,37	1,12	1,32
Nombre de composante	5	1	4
Taille de la grande composante (en % du total)	1198 (98,9)	997 (100)	1935 (99,6)
Nombre de liens dans la GC (en % du total)	26252 (99,9)	15462 (100)	40387 (99,99)
Distance moyenne* (DM)	2,37	2,42	2,44
Diamètre D	5	5	5
Coefficient de cluster (CCI)	0,860	0,849	0,851

¹Télécommunications ; ²Electronique-Microélectronique ; ³Les deux domaines confondus

* Distance moyenne dans plus grande composante.

Tableau 12 : Propriétés générales des réseaux associées à chaque domaine technologique et aux deux domaines confondus.

Nous examinons si ces réseaux sont des “small-world networks”. Nous commençons par les critères avancés par Goyal, van der Leij et Moraga-Gonzalez.

Premier critère, le réseau du domaine Electronique-Microélectronique forme une seule composante. Le réseau des Télécommunications est constitué de 5 composantes, la plus grande composante contenant la quasi totalité des participants (98,9%). Lorsque l'on observe le réseau agrégé, le nombre de composantes n'est plus que de 4 ce qui signifie que l'agrégation des deux réseaux permet d'obtenir un réseau encore plus intégré.

Second critère, on observe que la densité du graphe est assez faible ce qui indique que les agents possèdent peu de liens relativement à ce qu'ils pourraient réaliser. Toutefois, le degré moyen, mesuré ici dans le réseau binaire et qui correspond donc au nombre de collaborations distinctes de l'agent, est plus élevé que dans la plupart des études menées sur les réseaux réels.

Cependant la comparaison avec les résultats de ces études est délicate du fait de l'hypothèse retenue pour construire les réseaux.

Si l'on observe les propriétés du réseau agrégé, on remarque que la densité est moins forte que dans chaque réseau pris individuellement ce qui signifie que chaque domaine technologique constitue un espace plus intégré dans lesquels les collaborations sont plus nombreuses. Pour autant, on remarque que l'agrégation des deux domaines technologiques conduit aussi à un espace intégré. On remarque notamment que le diamètre du graphe est de 5 comme dans les réseaux de chaque domaine technologique et que la distance moyenne est faible. Le réseau agrégé n'a donc pas perdu en efficacité. Enfin, comme nous l'avons déjà remarqué, l'agrégation des deux réseaux a permis d'intégrer une composante du réseau Télécommunications à la plus grande composante. Les agents qui participent aux deux domaines technologiques assurent donc un rôle de relais important afin de permettre la création d'un espace intégré où il existe des connections entre les domaines technologiques.

Troisième critère, la distance moyenne et le diamètre du graphe sont très faibles et inférieurs à $\ln(n)$ pour tous les réseaux. Les agents peuvent donc échanger rapidement au sein du réseau tout en n'ayant qu'un nombre restreint de liens.

Quatrième critère : dans les trois réseaux le coefficient de cluster est élevé et proche de 1. Les partenaires d'un agent ont donc une forte probabilité d'être aussi partenaire les uns des autres. Ces coefficients de cluster élevé s'explique en grande partie par la manière dont nous avons construit les réseaux et notamment par l'hypothèse retenue au départ. Par la suite, à l'aide du coefficient de cluster proposé par Newman, Watts et Strogatz, nous montrerons qu'il faut relativiser ces valeurs.

Il s'avère donc que les trois premiers critères établis par Goyal, van der Leij et Moraga-Gonzalez pour définir un "small-world network" sont respectés par les réseaux de collaborations dans le cadre des PCRDT. Nous examinons maintenant si les réseaux satisfont aux critères énoncés par Newman, Watts et Strogatz.

	Réseau étudié (DM)	Réseau aléatoire (DM_{Alea})	<i>Ratio</i>
Tel*	2,374	1,935	1,23
EM**	2,42	2,193	1,11
Les deux domaines confondus	2,44	2,119	1,15

*Tel = Télécommunications ; **EM = Electronique-Microélectronique.

Tableau 13 : Comparaison de la distance moyenne des réseaux étudiées avec celui d'un réseau aléatoire.

Premier critère : dans les trois réseaux, la distance entre les agents est très faible et est proche de celle des réseaux aléatoires. De ce fait, même si les agents au sein des réseaux de collaboration des PCRDT sont situés à proximité les uns des autres, ils ne sont pas plus proches, malgré les incitations et la politique de l'UE, que si nous avions construit le réseau à partir d'un processus aléatoire.

Deuxième critère, les coefficients de cluster obtenus dans le tableau résumant les propriétés générales sont élevés ce qui est normal étant donné l'hypothèse de départ retenue pour construire le réseau. Pour contourner cette difficulté, nous utilisons l'indicateur proposé par Newman, Watts et Strogatz (2001). Le tableau ci-dessous donne le coefficient de cluster de chaque réseau, plus le coefficient de cluster pour un réseau aléatoire de même taille.

	Réseau étudié (C_{Clbis})	Réseau aléatoire (C_{ClAlea})	Ratio
Tel*	16,82 %	33,6 %	0,511
EM**	17,16 %	28,9%	0,582
Les deux domaines confondus	12,83 %	28,1 %	0,457

*Tel = Télécommunications ; **EM = Electronique-Microélectronique.

Tableau 14 : Comparaison du coefficient de cluster des réseaux étudiées avec celui d'un réseau aléatoire.

Pour les trois réseaux, on observe que le coefficient de cluster obtenu grâce à l'indicateur de Watts (1999) est assez faible. Lorsque l'on calcule le ratio entre le coefficient de cluster du réseau étudié et le coefficient de cluster du réseau aléatoire de taille identique, on observe que le ratio est proche de 0,5 pour les trois réseaux. Ceci signifie donc que les réseaux sont assez éloignés en termes de coefficient de cluster d'un "small world network". En effet, selon Watts le ratio doit excéder 1 et plus le ratio est élevé plus le réseau étudié se rapproche d'un "small world network". Les agents au sein des PCRDT ne semblent pas posséder assez de liens entre eux. Il semble donc qu'un faible nombre d'agents jouent le rôle d'intermédiaire ce qui expliquerait à la fois des distances entre agents et des coefficients de clusters faibles.

L'UE cherche à créer un EER comprenant un nombre important d'agents avec de nombreuses collaborations. C'est le cas pour les deux domaines technologiques étudiés. Ces deux domaines forment un espace intégré aussi bien individuellement que de manière agrégé. De plus, on peut voir que les distances au sein des réseaux sont faibles facilitant ainsi la transmission de connaissances même si l'on retient les hypothèses de "decay" et réseau borné.

Une partie des objectifs fixés par l’UE semble atteinte.

L’analyse réseau a ainsi permis de mettre en avant les principales propriétés des réseaux des deux domaines technologiques. Les réseaux formés par les collaborations au sein des PCRDT possèdent des caractéristiques propres aux “small world networks”. Seul le coefficient de cluster est faible (selon la mesure de Newman et al. (2001)). Les réseaux sont fortement connectés et forment des espace intégrés où quasiment tous les agents sont liés de manière directe ou indirecte. Chaque sous-domaine est un réseau particulièrement dense de collaborations mais qui n’est pas refermé sur lui même puisque l’on a vu qu’il existait des passerelles entre les domaines technologiques. Cela suppose donc qu’il existe des interactions entre domaines technologiques favorables à la production d’externalités technologiques, elles-mêmes favorables à l’innovation. L’analyse réseau a ainsi permis de mettre en avant que les réseaux de collaborations des PCRDT possèdent des propriétés favorables à l’innovation.

Toutefois, avant de conclure définitivement, il faut encore s’intéresser à la robustesse du réseau et à la distribution des liens pour voir s’il existe des “hubs” (des agents disposant d’un nombre important de liens) et si le réseau vérifie ainsi les autres propriétés associées aux “small world networks”.

2.1.2.3. Distribution des participations et des liens au sein des réseaux

Afin de compléter notre analyse des réseaux, nous nous intéressons à la distribution des participations et des liens pour chaque réseau. Deux raisons nous incitent à réaliser cette analyse : vérifier que les réseaux suivent des lois de puissance et déterminer si les réseaux contiennent des “hubs” au sens d’agents possédant de nombreux liens.

Distribution des liens et indice de Gini

Dans ce paragraphe, nous utilisons uniquement le graphe binaire. L’objectif est ici de déterminer s’il existe des “hubs”. Les hubs jouent un rôle important dans les “small world networks” en servant de relais aux autres agents. Le rôle de relais nécessite d’avoir des liens nombreux et variés, l’intensité du lien ne rentre pas en ligne de compte, seul l’existence de ce lien est intéressante.

Une première manière d’identifier s’il existe des “hubs” est de s’intéresser à l’indicateur de centralité du graphe. Dans la mesure où les “hubs” d’un

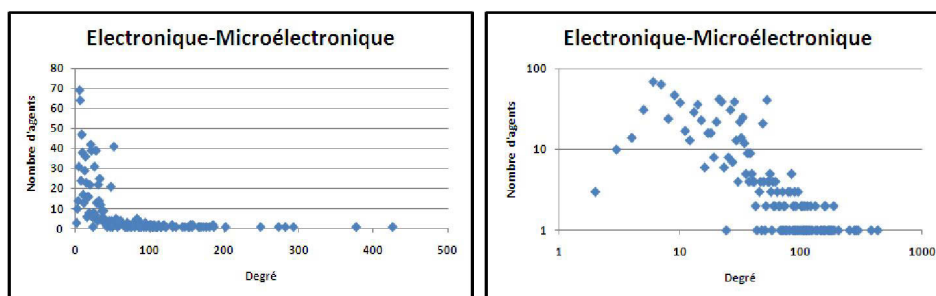
“Small world network” doivent posséder un nombre important de liens et qu’ils assurent une fonction de relais au sein du réseau, nous retenons comme indicateur de centralité le degré et le betweenness, tous deux calculés dans le réseau binaire. Les indicateurs obtenus pour les réseaux des deux domaines technologiques sont présentés dans le tableau ci-dessous :

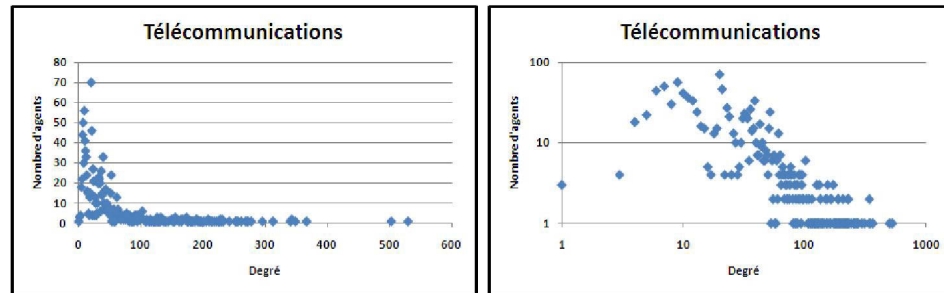
Réseaux	Degré	Betweenness
Electronique-Microélectronique	39,74%	16,85%
Télécommunications	40,20%	13,61%

Tableau 15 : Centralité du graphe.

Les mesures de centralité du graphe sont proches. La centralité de degré des graphes laisse supposer qu’il existe des agents qui possèdent un nombre important de liens. Pour autant, la mesure est loin de se rapprocher de celle obtenue dans le cas de réseau en forme d’étoile ce qui laisse supposer que plusieurs agents se partagent le rôle de “hubs”. Cette impression est confirmée par l’indicateur de betweenness des deux graphes qui est très faible. Il existe donc au sein du réseau un nombre important d’agents qui servent de relais. Afin de vérifier qu’il existe plusieurs “hubs” pour chaque réseau, nous construisons la distribution des liens.

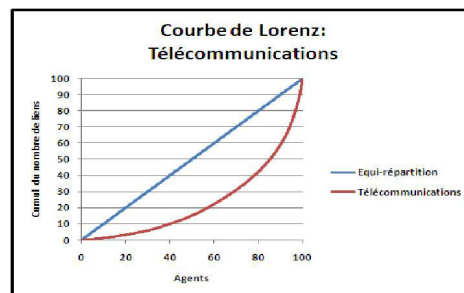
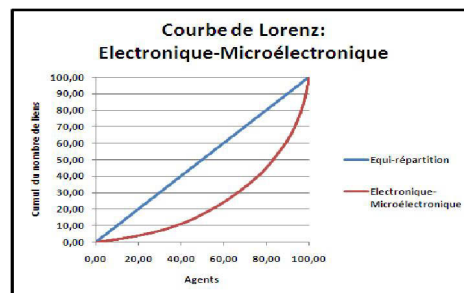
La distribution du nombre de liens permet d’observer l’allure du nuage de points formés par chaque distribution. De plus, afin d’observer si ces distributions suivent une loi de Poisson dans les graphiques ci-dessous, nous avons représenté cette distribution à l’aide de deux échelles : une échelle normale et une échelle logarithmique pour chaque axe.





Graphiques 15 à 18 : Distribution des liens des réseaux des deux domaines technologiques.

Lorsque l'on observe les graphiques, nous remarquons que si la distribution des liens de chaque réseau se rapproche de celle d'une loi de Poisson, elle n'en est pas une. Le graphique avec l'échelle logarithmique permet de voir que le nombre relativement important d'agents qui possèdent entre 1 et 10 liens ne permet pas de respecter la loi 80/20 de Pareto. Toutefois, il apparaît assez nettement que même si la loi des 80/20 n'est pas respectée, la répartition des liens est très inégale quel que soit le réseau. Pour vérifier cette assertion, nous représentons la courbe de Lorenz et nous calculons pour les deux réseaux l'indice de Gini. Les résultats sont présentés ci-après. Dans les deux cas, la courbe bleue représente le cas où les liens seraient répartis de manière équitable et la courbe rouge représente la distribution des liens du réseau étudié



Graphiques 19 à 20 : Répartition des liens entre agents et courbes de Lorenz

Réseaux	Indice de Gini
Electronique-Microélectronique	0,505
Télécommunications	0,532

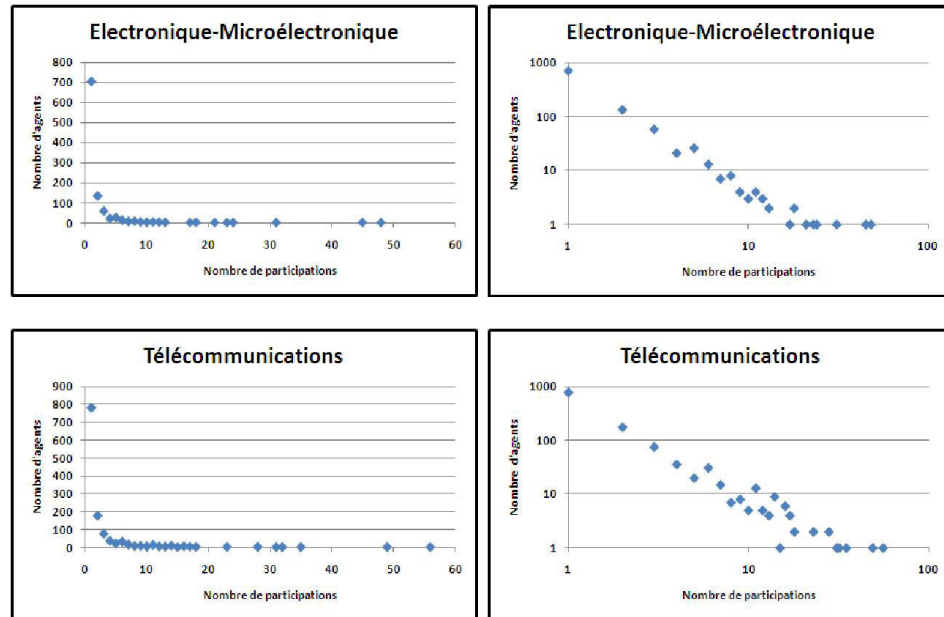
Tableau 16 : Indices de Gini de la répartition des liens des deux domaines technologiques.

Dans les deux réseaux, on observe une forte concentration des liens. Il existe donc, au sein des réseaux, des “hubs” qui possèdent un nombre de connexions très important. Ces “hubs”, par le nombre de liens qu’ils possèdent, sont des relais au sein du réseau. Par la suite, nous verrons si cette position de relais se confirme lorsque nous testerons la robustesse du réseau (cf. paragraphe 2.1.2).

La décision des agents de participer aux PCRDT dépend de nombreux paramètres (financement, volonté d’innover,...). L’un des paramètres les plus importants dans le cadre des PCRDT est d’intégrer un réseau de recherche et de nouer des collaborations. L’agent choisit de collaborer avec l’ensemble des agents du consortium en acceptant de participer à un projet. La décision de l’agent porte sur le fait de participer ou non à un projet si les agents qui le composent l’intéressent et s’ils souhaitent collaborer avec eux pour diverses raisons (les performances technologiques et économiques de ces agents, relation de confiance,...). Le nombre de projets par agent est une autre manière de vérifier la présence de “hubs”. On peut alors se demander si la distribution des participations (du nombre de projets par agents) ne suit pas une loi de Poisson et ainsi mettre en avant une des propriétés d’un “small-world network”.

Distribution des participations

De la même manière que l’on a construit pour chaque domaine technologique la distribution des liens au sein de leurs réseaux respectifs, nous construisons cette fois-ci la distribution des participations. Pour rappel, un agent qui participe à un projet, cela fait une participation. De fait, il est équivalent de parler de la distribution du nombre de participations ou du nombre de projets pour les agents des réseaux. Les distributions obtenues sont présentées dans les graphes ci-dessous.



Graphiques 21 à 24 : Distribution des participations des réseaux des deux domaines technologiques.

Sur les graphiques de gauche, on observe très clairement que les agents qui participent à un ou deux projets sont les plus représentés et ce pour les deux domaines technologiques. Le tableau ci-dessous donne le pourcentage d'agents qui participent au maximum à deux projets ainsi que le pourcentage de projets que cela représente.

Réseaux	Pourcentage d'agents à 1 ou 2 projets	Pourcentage de projets associés
Electronique-Microélectronique	84 %	49 %
Télécommunications	79 %	38 %

Tableau 17 : Agents comptant 1 ou 2 projet(s).

Les agents qui comptent 1 ou 2 participations représentent 80% des agents quel que soit le domaine technologique. Toutefois, dans les deux cas, le pourcentage de participations associé à ces agents dépasse les 20% de la règle de Pareto. Pourtant lorsqu'on observe les graphiques de droite utilisant une échelle logarithmique, les deux distributions semblent suivre une loi de Poisson. En effet, sur les graphiques de droite, de nombreux agents sont concentrés et il existe des agents à l'extrême droite de la distribution qui représentent peu d'agents mais dont le niveau de participation est très élevé.

Pour vérifier, si ces deux distributions suivent une loi de puissance, nous estimons la valeur de k pour chaque distribution à l'aide de la relation suivante :

$$\log(y) = k.\log(x) + \log(\alpha)$$

où (y) représente le nombre d'agents qui compte x participations. Les réseaux qui suivent une loi de Poisson sont aussi appelés "scale free network". Si les réseaux que nous étudions peuvent être comparés à un "scale-free network" c'est une preuve que nos réseaux sont des "small world networks". Cohen et Havlin (2003) ont montré que les "scale-free network" étaient caractérisés par une loi de puissance dont le paramètre k appartenait à l'intervalle $[-3; -2]$. Rappelons que si les réseaux sont des "Small world networks" alors la valeur du paramètre k doit être comprise entre -3 et -2. Après estimations, nous obtenons les résultats suivants pour le paramètre k .

Réseaux	Paramètre k
Electronique-Microélectronique	-1,776
Télécommunications	-1,706

Tableau 18 : Estimation des paramètres de la loi de Poisson pour les deux domaines technologiques.

Si on peut mettre en avant que le réseau suit une loi de puissance alors c'est une preuve que le réseau possède une des propriétés propres aux "small world-networks". Le paramètre de nos deux distributions est légèrement inférieur puisqu'il est proche de -1,7. La faiblesse du paramètre k s'explique en grande partie par le fait que quelques agents de la distribution possèdent des caractéristiques particulières. En effet, on remarque que seul un faible nombre d'agents possède un très grand nombre de participations. Ces agents, par leur poids en termes de nombre de participations, ont tendance à aplatiser la distribution. Il est possible d'utiliser un processus de "data binning" pour réduire l'effet de ces agents extrêmes sur la distribution. Pour cela, la procédure consiste à regrouper les agents par classe. A chaque classe correspond un poids et une valeur représentative permettant ainsi de lisser la distribution.

Au final, la distribution des liens et des participations, si elles ne permettent pas d'affirmer que les réseaux sont des "small world networks", laissent supposer l'existence de "hubs" qu'il faudra confirmer par la suite à l'aide des indicateurs de centralité (cf. sous-section 2.1.3).

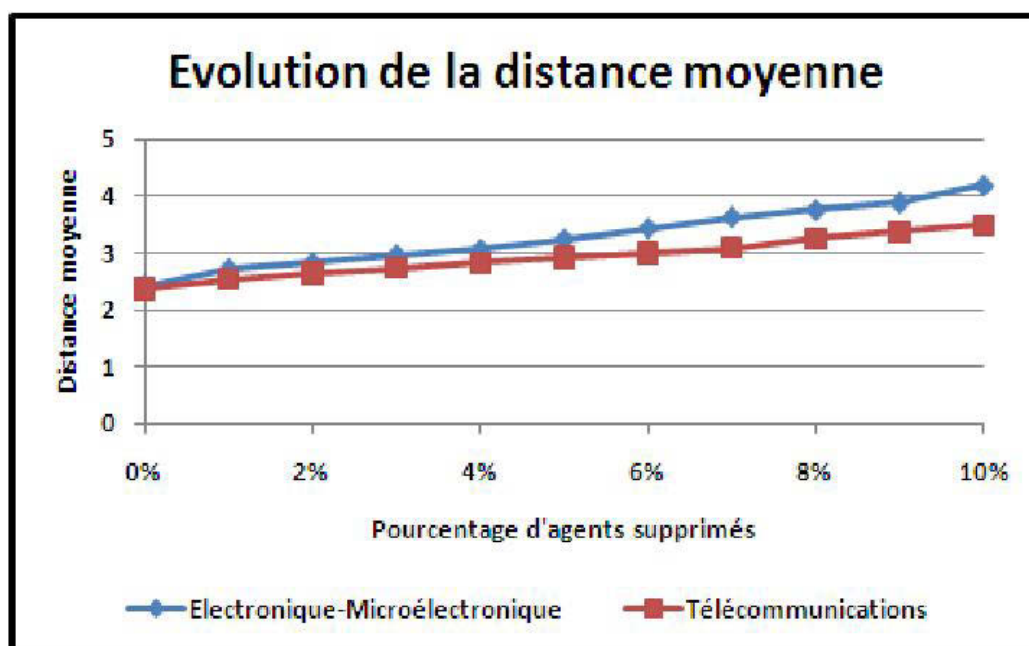
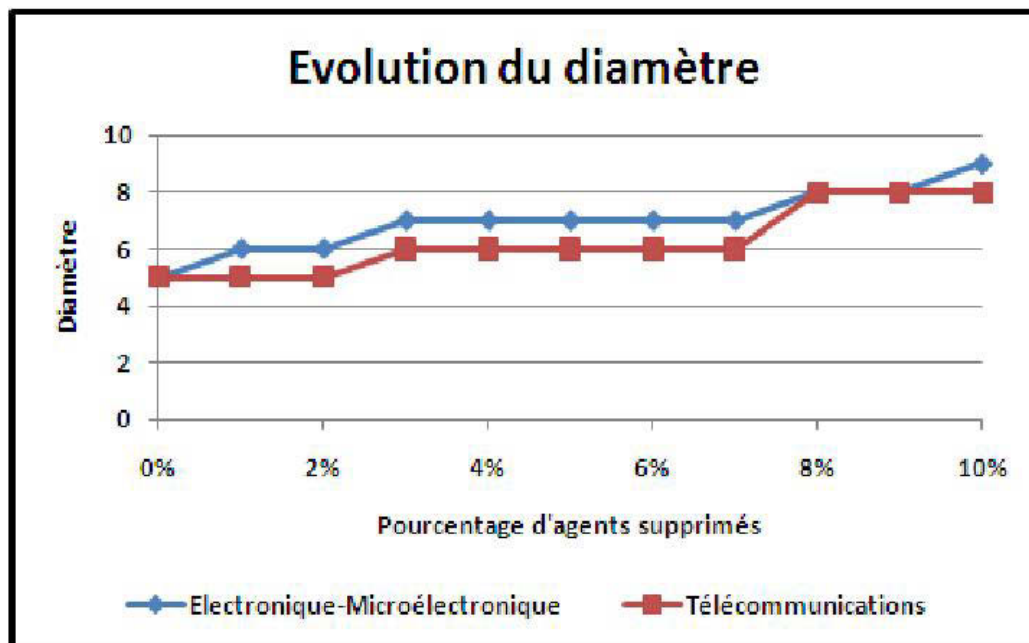
Dans le paragraphe suivant, nous observons si les réseaux des Télécommunications et de l'Electronique-Microélectronique sont robustes à la suppression des agents centraux. Rappelons que les "small world networks" sont

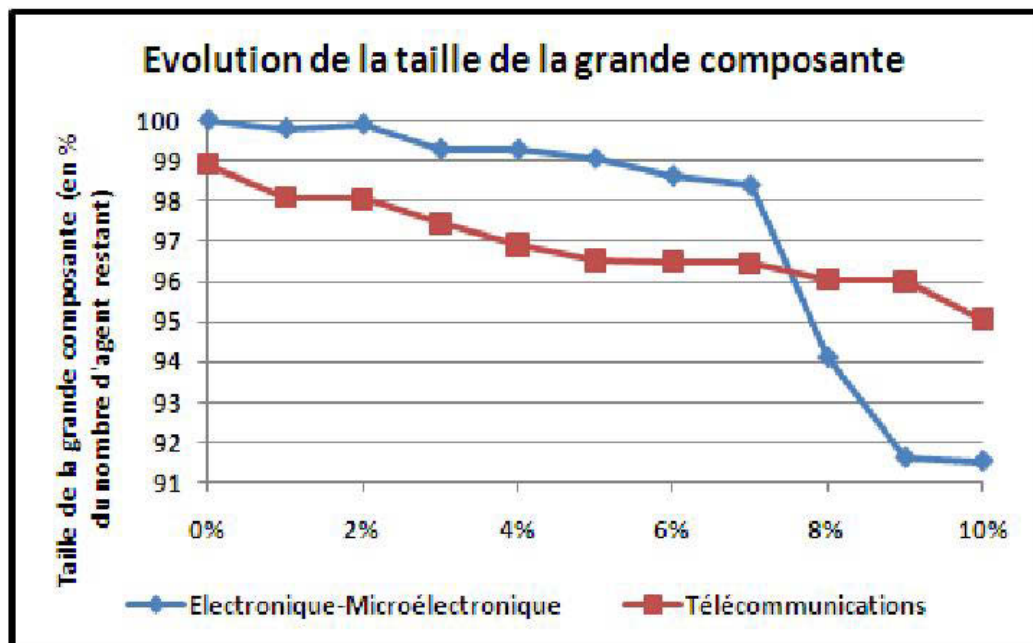
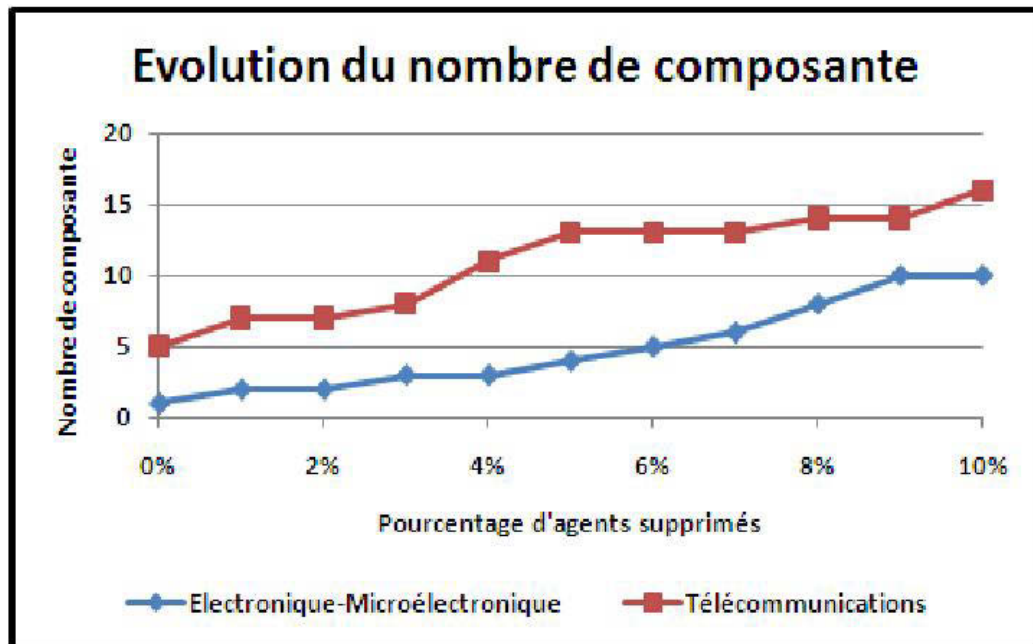
des réseaux supposés robustes à la suppression aléatoire d'agents mais par contre la suppression des agents les plus centraux peut provoquer de fortes modifications des propriétés du réseau et en altérer l'efficacité.

2.1.2.4. Robustesse des réseaux à la suppression des agents centraux

L'UE souhaite mettre en place un EER qui se veut un réseau de collaboration dense et intégré. Si cet espace doit reposer sur des agents centraux, il est important pour autant que le réseau de collaboration ne soit pas dépendant d'un faible nombre d'agents. L'apparition de "hubs" est nécessaire afin d'améliorer les propriétés générales du réseau et, de manière générale, il est important que des agents "leaders" émergent au sein des réseaux afin de tirer les projets d'innovation par le haut grâce à leurs connaissances, leurs moyens économiques, . . . Il est aussi important que le réseau ne dépende pas d'un faible nombre d'agents. L'UE est donc face à une situation où il est nécessaire d'avoir des "hubs" mais, en même temps, en avoir un certain nombre pour éviter de dépendre d'un ou deux agents. Dans le cas de notre étude, il faudrait donc que l'on arrive à montrer qu'il existe un nombre suffisant de "hubs" de manière à ce que le réseau ne soit pas trop affecté par la suppression d'un faible nombre d'entre eux. Dans l'idéal, il faudrait donc que les réseaux résistent à la suppression des premiers "hubs" et que les propriétés déclinent lorsque l'on commence à supprimer un nombre assez important d'agents.

Évolution des propriétés des réseaux suite à la suppression des "hubs" Pour évaluer l'impact des "hubs" sur le fonctionnement des réseaux et la robustesse de ces derniers, nous supprimons de manière séquentielle par tranche de 1% les "hubs" du réseau et leurs liens et nous évaluons à chaque fois l'impact de la suppression de ces agents sur les propriétés générales du réseau. Afin de respecter la définition de "hubs" dans le cas des "small world networks", nous supprimons pour chaque réseau les agents qui comptent le plus grand nombre de liens distincts. Nous retenons ici la mesure de degré appliqué au réseau binaire puisque la valeur des liens n'entre pas en compte dans le calcul de la plupart des propriétés générales. L'important pour nous est donc de supprimer les agents qui ont le plus de liens et non ceux qui ont le plus de collaborations. Les résultats complets sont présentés dans l'annexe A.2. Ci-dessous, nous visualisons l'impact de ces suppressions sur les propriétés suivantes : la distance moyenne, le diamètre, le nombre de composantes et la taille de la grande composante.





Graphiques 25 à 28 : Évolution des propriétés générales des réseaux suite à la suppression des agents les plus centraux. Les droites entre les points ont été rajoutées afin d'améliorer la lisibilité des graphiques. Les chiffres exacts sont disponibles dans l'annexe A.4.

La distance moyenne et le diamètre des deux réseaux évoluent de manière assez similaire. L'évolution de ces deux propriétés se fait de manière assez linéaire quel que soit le domaine étudié. En effet, on ne remarque aucun point de rupture significatif sur le graphique. Toutefois, l'accroissement du diamètre et de la distance moyenne est plus important concernant le domaine Electronique-Microélectronique. En effet, le diamètre passe de 5 à 9 lorsque celui du réseau Télécommunications augmente jusqu'à 8. Parallèlement, la distance moyenne du réseau Electronique-Microélectronique augmente de 73% lorsque celle du réseau Télécommunications augmente de seulement 46%. Le domaine Télécommunications semble donc légèrement plus robuste à la suppression des agents centraux que le réseau Electronique-Microélectronique mais dans les deux cas on note une augmentation sensible de distances qui peut être préjudiciable à la circulation des connaissances et des externalités.

Concernant le nombre de composantes, on observe un résultat plus mitigé. Le nombre de composantes est bien plus élevé dans le domaine des Télécommunications. Ce nombre passe de 5 à 16 pour le réseau des Télécommunications alors que le réseau compte 10 composantes suite à la suppression des 10% d'agents les plus connectés contre une seule au départ dans le domaine Electronique-Microélectronique. De manière relative, l'augmentation est bien plus importante dans le cas du réseau Electronique-Microélectronique. Le réseau Télécommunications apparaît donc plus éclaté, mais, si l'on observe le nombre d'agents restants dans la grande composante, on s'aperçoit qu'il existe un point de rupture tel que le nombre d'agents dans la grande composante devient plus important dans le domaine des Télécommunications. Le nombre d'agents restant dans la grande composante du réseau Electronique-Microélectronique diminue assez fortement. A ce seuil, on observe que la suppression d'un agent a causé l'apparition d'une composante de 36 agents. Dans le réseau Télécommunications la taille de la deuxième plus grande composante n'excède pas 9 agents. Si le nombre de composantes est plus élevé dans le réseau Télécommunications, ce dernier conserve une grande composante plus imposante. Dans les deux cas, il faut remarquer que la grande composante compte plus de 90% des agents restants du réseau. Si l'on faisait l'hypothèse que les informations circulent de manière parfaite au sein d'une composante du réseau et que la distance entre les agents importe peu, le fait de conserver une composante comptant presque tous les agents du réseau permet de conserver presque intactes les capacités du réseau à faire circuler l'information.

Au final, la suppression des 10% des agents comptant le plus grand nombre de liens affecte les propriétés des réseaux. Les distances sont presque doublées et le nombre de composante augmente. Toutefois, la plus grande

composante résiste assez bien à la suppression des agents les plus connectés puisqu’au final, pour chaque domaine technologique, plus de 90% des agents restants en font encore partie. Il ne faut pas omettre que 10% des agents supprimés dans les réseaux représentent plus de 100 agents au total. On montre donc que les réseaux possèdent un nombre important de “hubs” tel que la suppression des “hubs” comptant le plus de liens n’affecte pas de manière trop importante les propriétés générales du réseau tant que le nombre d’agents supprimés n’excède pas 5 à 6 % selon les indicateurs. De cette manière, nous avons montré que les réseaux possèdent des “hubs” nécessaires au fonctionnement du réseau mais en assez grand nombre pour ne pas dépendre trop de chacun d’entre eux.

Selon les hypothèses retenues concernant la manière dont circulent les informations au sein du réseau, cette situation peut être vue de manière différente. Dans la mesure où de nombreux agents participent à plusieurs projets (environ 30% dans les deux domaines technologiques), il apparaît probable qu’il est important de participer activement aux PCRDT et d’être bien placé au sein du réseau pour accéder à un maximum de connaissances, d’informations et de bénéficier au mieux des externalités. La présence au sein du réseau ne semble pas suffisante pour assurer une parfaite captation des connaissances qui circulent. Si c’était le cas, le nombre d’agents comptant plus d’une participation serait plus faible. En outre, il semble important au sein du réseau des PCRDT de collaborer avec des agents influents. En effet, au vu du nombre important d’agents possédant de nombreux liens, il apparaît fortement probable que les agents identifiés comme “hubs” possèdent de nombreux liens entre eux. Nous montrerons dans la partie II que le choix des collaborations est influencé par le prestige des agents et que la centralité des agents peut jouer un rôle important.

Globalement, les réseaux construits à partir des collaborations aux PCRDT possèdent des propriétés proches de celles des “small world networks”, propriétés qui sont favorables à la diffusion de connaissances de manière efficace. Les objectifs de l’UE, qui cherchaient à créer au travers des PCRDT, un EER sont partiellement atteints pour les deux domaines technologiques étudiés. Les réseaux forment un espace intégré dans lequel de nombreux agents collaborent. Les principales composantes de chaque réseau contiennent tous les agents ou presque. Les distances entre les agents sont faibles et le diamètre des réseaux peu élevé facilitent la diffusion des connaissances et des externalités. De plus, il existe des passerelles entre domaines technologiques représentées par les agents qui réalisent au moins une participation dans chaque domaine technologique. Ces agents assurent ainsi que chaque domaine technologique

ne soit pas un espace cloisonné mais qu'il existe des interactions entre des domaines proches. Nous avons même pu remarquer que l'agrégation des deux réseaux diminuait le nombre de composantes et que pour autant l'efficacité du réseau n'était pas remise en cause puisque la distance moyenne et le diamètre n'augmentaient pas.

Il reste à déterminer quelles sont les caractéristiques de ces agents. Nous avons établi que ces réseaux de PCRDT possèdent des "hubs" dont ils sont en partie dépendant. En effet, nous avons montré que suite à leur suppression les propriétés du réseau étaient altérées. Toutefois, il faut relativiser cette dépendance dans la mesure où les réseaux ne dépendent pas que de quelques "hubs". En effet, il faut supprimer 10% des agents pour que les propriétés générales du réseau soient altérées de manière importante soit 121 agents dans le cas du réseau Télécommunications et 100 dans le réseau Electronique-Microélectronique.

Les conclusions que nous obtenons ne sont valables que pour les deux domaines technologiques étudiés. La comparaison entre les deux domaines technologiques a mis en avant des caractéristiques similaires entre les deux réseaux, en plus des interactions entre eux. Les deux réseaux, bien que de taille différente, présentent des propriétés relativement similaires (distance moyenne, diamètre, densité, ...). De même, les distributions du nombre de participations par agent et du nombre de projet par agent sont semblables comme le montre l'estimation du paramètre k dans le cas des distributions du nombre de participation. Pour pouvoir réellement établir que les réseaux générés par les PCRDT forment un espace intégré, il faudrait étudier l'ensemble des thématiques technologiques. Il serait alors intéressant de voir s'il existe des agents qui participent à différentes thématiques technologiques et si nous aboutissons au final à une seule composante.

2.1.3 Des agents aux rôles différents

Jusqu'à présent les "hubs" ont été identifiés comme les agents possédant le plus grand nombre de liens distincts. Il s'agit ici d'une définition relativement restrictive. Nous montrerons que les indicateurs de centralité peuvent permettre d'affiner l'identification des "hubs" afin de déterminer qui sont les agents les plus centraux des réseaux. De plus, nous observerons aussi les différences entre les types d'agents et nous étudierons plus particulièrement le rôle joué par les PME. Enfin, nous accorderons une importance particulière aux agents qui permettent de créer des passerelles entre les deux domaines technologiques.

Identification des “hubs” ou de manière plus générale des agents centraux

Nous avons vu dans la section 1.4 que la centralité renvoyait à plusieurs définitions, chacune associée à un indicateur de centralité. Selon la définition retenue, les agents centraux ne seront pas forcément identiques. Nous cherchons à déterminer quels sont les agents les plus importants pour le réseau de chaque domaine technologique. Contrairement à la définition la plus courante d’un “hubs”, le nombre de liens n’est qu’un des critères. De fait, pour être qualifiés de “hubs”, les agents doivent réunir plusieurs qualités qui correspondent chacune à un indicateur de centralité. Cet agent doit :

- Participer à un grand nombre de projets ;
- Collaborer avec de nombreux partenaires distincts (degré dans le réseau binaire) ;
- Être un partenaire privilégié des agents les plus centraux ou avoir des partenaires communs (eigenvector centrality et/ou degré renforcé dans le réseau valué) ;
- Être le plus proche possible de l’ensemble des agents afin de bénéficier des connaissances des autres agents et des externalités du réseau (centralité de proximité) ;
- Être un relais au sein du réseau afin d’avoir un certain pouvoir sur le fonctionnement du réseau et capter les informations et connaissances (centralité d’intermédiation).

Au travers de la recherche des “hubs” des réseaux, il est aussi possible de chercher à vérifier un certain nombre d’objectifs fixés aux PCRDT. L’un des objectifs de l’UE était d’attirer dans les réseaux de collaborations les agents leaders dans leur domaine technologique. Dans le cas des entreprises, les agents leaders sont des groupes mondiaux soit des leaders nationaux. Pour les agents de l’enseignement supérieur ce sont plutôt des grandes universités ou les centres de recherche nationaux. Si pour chaque domaine technologique sont présents les agents les plus compétents, alors cela peut générer des effets d’externalités positifs. Cela permet aux autres agents participant aux PCRDT de profiter des connaissances et des compétences de ces agents. De plus, cela assoit la crédibilité de la politique des PCRDT. Si les agents leaders dans chaque domaine sont présents dans les PCRDT, nous pouvons supposer qu’ils tirent profit des collaborations qu’ils nouent au sein des programmes cadres. Étant donné que le coût supporté par la plupart des agents est supérieur au financement qu’ils obtiennent, leur participation implique que les agents considèrent la collaboration comme bénéfique pour leur activité de R&D présente et future.

La difficulté d'un classement des agents en termes de centralité réside dans le fait d'une part que les agents n'occupent pas le même rang pour les différents indicateurs, d'autre part qu'il n'est pas possible de comparer les mesures obtenues par un agent pour deux indicateurs différents. La seule comparaison possible entre deux mesures est le rang de l'agent (ou son classement). Il est possible d'établir un classement moyen sur les 5 critères de centralité sans pondération (aucun des critères proposés n'a plus de valeur qu'un autre). De cette manière, nous obtiendrons un classement moyen de l'agent. L'ordre ainsi établi permettra de dégager un sous-ensemble d'agents qui satisfont le mieux possible les 5 critères précédemment présentés. Pour obtenir le rang moyen d'un agent i ($Av.rank(i)$), nous retenons la moyenne harmonique (Zimmermann, 2007) qui se calcule de la manière suivante :

$$Av.rank(i) = |I_C| \times \frac{1}{\sum_{k=1}^{|I_C|} \frac{1}{r_k(i)}}$$

avec I_C l'ensemble des indicateurs de centralité retenus comme critère, $|I_C|$ le nombre d'indicateurs, $k \in I_C$ un indicateur particulier et $r_k(i)$ le rang de l'agent i pour l'indicateur k .

Nous avons retenu une moyenne harmonique plutôt qu'arithmétique pour deux raisons. La première, c'est que la moyenne harmonique permet de récompenser ceux qui sont très bien classés dans la mesure où, par exemple, être premier compte deux fois plus qu'être second. Parallèlement, la différence de rang entre deux agents classés l'un derrière l'autre tend vers 0 lorsque le classement de ces agents tend vers l'infini. C'est un avantage puisque certains indicateurs, notamment le degré et le nombre de projets, sont peu discriminants. De fait, de nombreux agents se retrouvent avec des niveaux de centralité proches et plus particulièrement dans le bas du classement. Par exemple, un seul projet de différence peut faire évoluer le classement de manière importante. La moyenne harmonique permet d'éviter que de faibles écarts entre les agents se traduisent par de trop fortes différences sur le classement moyen.

La deuxième raison, c'est que la moyenne harmonique a l'avantage de ne pas trop pénaliser les agents qui ne sont faibles que dans un critère contrairement à la moyenne arithmétique. Dans les différents cas où on utilise le concept de rang moyen ou de classement moyen, on utilise souvent plus de 5 critères. Généralement, pour éviter d'introduire dans la moyenne des "accidents", on enlève la contribution amenée par le meilleur et le moins bon classement d'un agent. Dans notre cas, ce serait délicat car nous ne possédons que 5 critères. La moyenne harmonique permet de limiter le poids d'un seul mauvais classement pour l'agent. En revanche, elle a tendance à mettre

trop en avant les bonnes performances.

Le classement obtenu pour les deux domaines technologiques est présenté en annexe (cf. Annexe A.5). Seuls les 40 agents les plus centraux de chaque domaine technologique apparaissent dans l'annexe.

Si on observe le domaine des Télécommunications, on remarque que parmi les agents les plus centraux figurent de nombreux groupes de téléphonie aussi bien des constructeurs d'appareil mobile comme Nokia ou Ericsson mais aussi des opérateurs comme France Telecom ou Telecom Italia. Ces agents sont soit des leaders nationaux comme France Telecom soit des groupes mondiaux comme Nokia. Dans le domaine Electronique-Microélectronique, le groupe ST Microelectronics et le groupe Philips comptent deux de leur établissements/filiales parmi les 40 agents les plus centraux de ce réseau. Par exemple, le groupe ST Microelectronics emploie 5000 personnes et en 2008 le montant total des ventes du groupe a presque atteint 10 milliards de dollars (Source :<http://www.st.com/stonline/>). ST Microelectronics fait partie des leaders dans la production de semi-conducteurs. Philips est l'un des plus grands groupes mondiaux en matière d'électronique. Quant au groupe Philips, il compte 15 entreprises/filiales qui participent aux deux domaines technologiques et totalisent 101 participations au total. Thales (anciennement Thomson-CSF), société d'électronique spécialisée dans l'aérospatial, la défense et les technologies de l'information participe quand à elle aux deux domaines technologiques par l'intermédiaire de filiales différentes. Dans les deux domaines technologiques, on retrouve parmi les agents les plus centraux des grands groupes industriels.

Ces grands groupes industriels ne constituent toutefois pas la majorité des agents les plus centraux. En effet, aussi bien dans le domaine Electronique-Microélectronique que dans le domaine des Télécommunications, on retrouve en majorité parmi les 40 agents les plus centraux des agents de l'enseignement supérieur (des universités). Parmi ces agents de l'enseignement supérieur, certains sont classés dans les deux domaines technologiques comme l'Ecole Polytechnique de Lausanne. Enfin, ce classement est complété par les centres de recherche. En France, dans le domaine Electronique-Microélectronique, le Centre National de la Recherche Scientifique (CNRS) et le Commissariat à l'Énergie Atomique (CEA) qui font partie des grands centres de recherche français sont parmi les agents les plus centraux. L'agent le plus central pour chaque domaine technologique est identique, il s'agit du centre de recherche Fraunhofer Gesellschaft zur Foerderung Angewandten Forschung e.v. Ce centre est le premier en termes de nombre de projets et de nombre de liens distincts quel que soit le domaine technologique. Il comptabilise 101 participations au total (56 dans le domaine des Télécommunications et 45

dans le domaine Electronique-Microélectronique). De plus, dans les réseaux, il est situé à proximité des autres agents et jouent un rôle de relais important comme en témoigne son indicateur de betweenness. Le seul critère où il est devancé est l'“eigenvector centrality”. Dans le domaine Electronique-Microélectronique, c'est le centre de recherche IMEC (centre de dimension européenne employant 1650 personnes) qui est premier pour cet indicateur et, dans le réseau Télécommunications, il s'agit du groupe Espagnol Telefonica.

En résumé, l'objectif de l'UE n'est pas seulement de voir participer des agents importants mais aussi de les voir jouer un rôle actif au sein des réseaux de collaborations. De nombreux grands groupes industriels au niveau mondial ou national font partie des agents les plus centraux ce qui témoigne de leur intérêt pour les activités de R&D qui se déroulent au sein des PCRDT. La participation active de ces agents améliore le niveau de la recherche et attire par la suite de nouveaux partenaires. En plus de ces grands groupes industriels, on remarque la présence de grands centres de recherche comme l'IMEC ou le CNRS ou encore des grands universités ou écoles européennes. L'objectif que s'est fixé l'UE semble donc atteint du moins pour les deux domaines technologiques étudiés. Malerba, Vonortas, Breschi et Cassi (2006) étaient arrivés à des conclusions similaires.

On peut cependant faire plusieurs remarques qui viennent nuancer cette affirmation.

On constate l'absence parmi les agents centraux de deux types d'agents que l'UE souhaitait favoriser. La première catégorie concerne les agents localisés dans les nouveaux États membres. Peu d'agents localisés dans ces États font partie de ces classements. Dans les deux domaines figurent seulement quatre agents localisés dans les nouveaux états membres, ce sont à chaque fois des universités. Parmi ces quatre agents, deux participent aux deux domaines technologiques. Il s'agit de l'Université technologique de Varsovie et de celle de Budapest. D'une manière générale, les agents les plus centraux sont localisés essentiellement dans les pays de l'UE des 15 avec une forte part d'agents issus du Benelux et des pays nordiques. Il apparaît normal que peu d'agents issus de ces nouveaux États membres fassent partie des agents centraux car le 6^{ème} PCRDT se déroulait pendant le processus d'intégration de ces pays. Pour autant, l'absence ou le faible nombre d'agents centraux ne reflète pas forcément le degré d'insertion de ces nouveaux États membres. Il serait intéressant d'étudier de quelle manière ces pays s'intègrent aux réseaux européens puisqu'il s'agit d'un des objectifs de l'UE. A un niveau individuel, il est difficile d'observer de quelle manière ces pays s'intègrent au réseau européen. Il est alors préférable de travailler à un niveau plus agrégé comme celui des régions européennes. Ce sera l'objet de la section 2.2.

La seconde catégorie d'agents que l'UE souhaitait mettre en avant au sein des PCRDT, ce sont les PME. Quel que soit le domaine technologique étudié, aucune PME ne figure parmi les agents les plus centraux. Cela s'explique en partie par le fait qu'une PME est par définition une structure de faible taille avec des moyens limités qui peuvent l'empêcher de participer de manière intensive aux PCRDT. De fait, les PME participent à peu de projets. Même si les PME ne font pas partie des agents centraux, il serait intéressant de voir quel rôle elles jouent au sein des réseaux. De manière plus générale, il serait intéressant d'examiner si chaque type d'agent joue un rôle différent dans le fonctionnement du réseau. Pour cela, nous pouvons calculer pour chaque type d'agent et pour chaque indicateur de centralité, un indicateur moyen qui reflèterait le rôle joué par chaque type d'agent. Ce sera l'objet du paragraphe suivant.

La dernière remarque est le constat que peu d'agents figurent simultanément dans le classement des 40 agents les plus centraux. En effet, sur les 40 agents les plus centraux du réseau Télécommunications, seuls 5 d'entre eux font partie des agents les plus centraux du réseau Electronique-Microélectronique. Si l'on s'intéresse seulement aux 20 premiers agents, un seul agent fait partie des deux classements. Dans le même temps, sur les 40 agents les plus centraux du réseau Télécommunications, 37 participent aux deux domaines technologiques et 35 dans le domaine Electronique-Microélectronique. Les agents, même s'ils appartiennent aux deux réseaux, ont tendance à concentrer leur moyen sur un domaine en particulier. Ceci peut être expliqué par le fait que l'un des deux domaines constitue leur cœur de métier comme le domaine Télécommunications pour les opérateurs de Téléphonie par exemple mais que dans le même temps, ils ont intérêt à participer au second domaine, domaine qui peut être utile à leur activité (par exemple fabrication des composants électronique pour les téléphones portables, nouvelles technologies comme les nanotechnologies, ...). Ces agents se voient dans l'obligation de partager aussi leurs ressources et ne peuvent donc pas participer autant à deux domaines technologiques.

De la même manière que l'on identifie les agents centraux de chaque domaine technologique, il est intéressant d'identifier les agents centraux du réseau agrégé. Les résultats sont présentés dans l'annexe A.6. Sur ces 40 agents, 39 participent aux deux domaines technologiques. Seul Telenor une entreprise de télécommunication basée en Norvège ne participe qu'au domaine des Télécommunications mais figure pourtant parmi les agents les plus centraux. Parmi ces 40 agents, 32 étaient déjà présents dans le classement des agents les plus centraux du réseau Télécommunications. Ceci s'explique par le fait que le réseau des Télécommunications compte un plus grand nombre de pro-

jets et d'agents. Sur ces 40 agents, on dénombre 23 agents de l'enseignement supérieur, 11 entreprises et 6 centres de recherche. Le point commun entre ces 40 agents est qu'ils possèdent tous un rayonnement européen voire mondial. Enfin, de manière logique, le centre de recherche Fraunhofer Gesellschaft est l'agent le plus central du réseau agrégé quel que soit le critère utilisé. Cet agent se détache nettement au niveau des collaborations. Il participe à 101 projets sur 354 et compte 854 partenaires distincts soit près de 45% des agents du réseau.

La combinaison des différents indicateurs de centralité a permis de mettre en avant les agents les plus centraux de chaque domaine technologique. L'analyse réseau par l'intermédiaire de ces différents indicateurs permet de mettre en évidence des agents aux priorités particulières et ne prend pas seulement en compte le nombre de liens des agents. Dans un contexte comme celui des PCRDT, cela permet de vérifier s'il y a adéquation entre ce que l'on peut entendre par agent important en dehors des réseaux et agents centraux. Nous avons pu remarquer que l'UE a réussi à attirer au sein de chaque réseau des agents leaders dans leur domaine, ce qui renforce la crédibilité et l'attractivité des réseaux européens même s'il serait nécessaire de disposer de données supplémentaires, par exemple relatives aux marchés couverts par ces domaines technologiques, pour déterminer si l'UE a réussi à attirer les plus grandes entreprises concernées par le domaine. On constate que les entreprises ne sont pas les agents les plus nombreux parmi les agents centraux contrairement aux agents de l'enseignement supérieur. De plus, aucune PME ne figure parmi les agents les plus centraux ni aucun agent de type autres. Il semble donc que tous les agents ne jouent pas un rôle identique au sein de ces réseaux.

Des disparités entre les différents types d'agents

L'UE souhaite faire participer différents types d'agents aux PCRDT. L'intérêt est d'inciter des agents, qui ont des objectifs différents et une manière différente de concevoir la recherche, à collaborer ensemble. Ainsi, en schématisant, on peut trouver d'un côté les agents de l'enseignement supérieur et les centres de recherche qui mènent une recherche fondamentale ou académique et de l'autre les entreprises qui mènent une recherche plus appliquée afin d'aboutir à un produit, service ou procédé commercialisable dans un but lucratif. Par exemple, l'étude de A.B. Jaffe (1989), confirmée depuis par de nombreuses autres études menées dans le contexte européen (cf. Autant-Bernard, Billand et Massard, 2010 pour une revue de littérature) a montré l'intérêt que des agents différents pouvaient avoir à collaborer. De la même

manière, il est important pour l'UE d'impliquer les PME qui n'ont pas forcément la possibilité d'intégrer des réseaux de collaborations de grande envergure. En effet, dans sa volonté de construction d'un EER, l'UE considère les PME comme des agents innovants susceptibles d'apporter une réelle plus value à la recherche européenne. Il reste à déterminer dans quelle mesure ces différents agents s'intègrent au réseau et le rôle qu'ils y jouent.

Pour rappel, l'UE distingue cinq types d'agent différents : les agents de l'enseignement supérieur (HE), les entreprises non PME (IND), les PME (IND PME), les centres de recherche (RES) et les agents de type autres (OTH). Les comportements des différents types d'agents sont assez hétérogènes notamment en termes de participations comme le montrent les deux tableaux ci-dessous.

Electronique-Microélectronique										
Nombre de projets	HE		IND		IND PME		OTH		RES	
	NA*	R**	NA	R	NA	R	NA	R	NA	R
1	124	49,01	170	71,43	246	88,49	53	81,54	112	68,71
2	45	17,79	32	13,45	26	9,35	11	16,92	20	12,27
3	28	11,07	16	6,72	5	1,80	1	1,54	8	4,91
4	11	4,35	4	1,68	0	0,00	0	0,00	6	3,68
5	18	7,11	4	1,68	1	0,36	0	0,00	3	1,84
6 et plus	27	10,67	12	5,04	0	0,00	0	0,00	14	8,59
Total	253	100	238	100,00	278	100,00	65	100,00	163	100,00

(* NA = Nombre d'agents ; ** R = Répartition en %)

Télécommunications										
Nombre de projets	HE		IND		IND PME		OTH		RES	
	NA*	R*	NA	R	NA	R	NA	R	NA	R
1	111	43,02	220	63,04	219	85,21	140	82,35	93	52,54
2	41	15,89	52	14,90	33	12,84	20	11,76	31	17,51
3	29	11,24	27	7,74	0	0,00	5	2,94	15	8,47
4	15	5,81	10	2,87	1	0,39	2	1,18	8	4,52
5	9	3,49	4	1,15	2	0,78	1	0,59	4	2,26
6 et plus	53	20,54	36	10,32	2	0,78	2	1,18	26	14,69
Total	258	100,00	349	100,00	257	100,00	170	100,00	177	100,00

(* NA = Nombre d'agents ; ** R = Répartition en %)

Tableaux 19 et 20 : Distribution du nombre de participations (ou de projets) pour chaque type d'agents.

Ces deux tableaux montrent que les agents de l'enseignement supérieur et, à un degré moindre, les centres de recherche sont les agents les plus participatifs. Dans le domaine Electronique-Microélectronique, 50% des agents de l'enseignement supérieur comptent au moins deux projets. Ce chiffre est encore plus important dans le domaine des Télécommunications où 57% des agents de l'enseignement supérieur comptent au moins deux projets. Les entreprises non PME sont situées à un niveau intermédiaire. La participation à plusieurs projets confère à ces agents un rôle d'intermédiaire au sein du réseau. Les PME et les agents de la catégorie Autres sont des agents mono projet pour la plupart. Quel que soit le domaine technologique observé, plus de 80% de ces agents ne participent qu'à un seul projet. En ne participant qu'à un seul projet dans la plupart des cas, les PME ne jouent pas un rôle d'intermédiaire. Notons toutefois qu'elles n'ont pas vocation à devenir des agents centraux du réseau. Leur participation au réseau leur assure déjà de pouvoir bénéficier des connaissances, d'une part, par leurs collaborateurs directs et, d'autre part, par leurs liens indirects au sein du réseau. Elles peuvent ainsi bénéficier des externalités de connaissances produites par les autres collaborations.

Afin d'analyser un peu plus en détail le rôle de chaque type d'agent et, plus particulièrement, celui des PME, nous calculons pour chaque type d'agent un indicateur de centralité moyen. Ainsi, le degré moyen du type PME est égal à la moyenne du degré de toutes les PME du réseau. Les indicateurs retenus sont les mêmes que ceux utilisés pour identifier les agents centraux. Nous calculons un indicateur moyen pour chaque type d'agent au sein de chaque réseau. Cet indicateur moyen est calculé à partir des indicateurs normalisés. Les résultats sont présentés dans les tableaux ci-dessous.

	Electronique-Microélectronique				
Type 1	Proj. Norm.*	CD	CC	CB	CE
HE moyenne (EC**)	1,65 (1,80)	4,91 (4,58)	44,46 (5,05)	0,25 (0,65)	3,56 (4,37)
IND moyenne (EC)	1,11 (1,47)	2,61 (2,65)	41,71 (4,41)	0,08 (0,28)	1,94 (3,23)
IND PME moyenne (EC)	0,68 (0,27)	1,63 (1,24)	39,71 (3,58)	0,01 (0,07)	0,94 (1,03)
OTH moyenne (EC)	0,72 (0,26)	1,83 (1,56)	39,39 (4,42)	0,01 (0,04)	0,93 (1,05)
RES moyenne (EC)	1,59 (3,52)	4,11 (5,71)	43,01 (5,20)	0,35 (1,75)	2,97 (6,34)
Moyenne générale (EC)	1,18 (1,88)	3,11 (3,82)	41,91 (4,89)	0,14 (0,80)	2,17 (3,92)

(* Proj. Norm. = Mesure du nombre de projet normalisé ; ** EC = Ecart-type)

	Télécommunications				
Type 1	Proj. Norm.*	CD	CC	CB	CE
HE Moyenne (EC**)	2,10 (2,65)	6,07 (6,02)	45,86 (5,02)	0,25 (0,61)	3,44 (4,97)
IND PME Moyenne (EC)	1,40 (2,41)	3,82 (4,74)	43,23 (4,98)	0,11 (0,58)	1,99 (4,11)
IND PME Moyenne (EC)	0,65 (0,37)	1,70 (1,43)	40,12 (4,03)	0,00 (0,02)	0,50 (0,81)
OTH Moyenne (EC)	0,71 (0,50)	1,90 (1,54)	39,75 (4,96)	0,02 (0,08)	0,49 (0,85)
RES Moyenne (EC)	1,59 (2,63)	4,31 (4,84)	44,23 (4,73)	0,18 (1,08)	2,12 (3,57)
Moyenne générale (EC)	1,32 (2,13)	3,66 (4,58)	42,80 (5,27)	0,11 (0,60)	1,80 (3,67)

(*Proj. Norm. = Mesure du nombre de projet normalisé ; **EC = Ecart-type)

Tableaux 21 et 22 : Indicateurs de centralité moyen pour chaque type d'agents.

Ces deux tableaux confirment les résultats observés précédemment. Les agents de l'enseignement supérieur (HE) et les centres de recherche (RES) sont les agents les plus participatifs quel que soit le domaine technologique

étudié. Par leurs nombreuses participations et le nombre important de liens qu'ils créent, ils sont ceux qui sont situés le plus à proximité des autres agents et servent de relais. En effet, on remarque que ces deux types d'agents sont ceux qui ont la mesure de *betweenness* la plus élevée. Ils sont donc les mieux placés pour profiter des connaissances et des externalités qui circulent au sein du réseau. Étant donné qu'ils possèdent l'indicateur d'eigenvector centrality le plus élevé et qu'ils constituent une grande partie des agents centraux, on peut en déduire que les centres de recherche et les agents de l'enseignement supérieur collaborent ensemble et que ces collaborations sont plus intenses. Nous montrerons notamment cela dans la partie 2 de la thèse.

Les entreprises non PME sont situées plus en retrait en termes de participations. Ce type d'agent possède des indicateurs de centralité proches de ceux d'autres agents dans le réseau (mesure de *closeness*) comme les agents de l'enseignement supérieur et les centres de recherche mais sans pour autant jouer un rôle d'intermédiaire comme le montre leur faible mesure de *betweenness*. Les PME et les agents de la catégorie autres participent peu comparativement aux autres agents (souvent à un seul projet). Ce faible niveau de participation impacte par la suite tous les autres indicateurs. Le *betweenness* de ces agents est proche de 0 ce qui signifie qu'ils ne jouent jamais le rôle d'intermédiaire. L'importance des PME est encore plus limitée dans le réseau des Télécommunications que dans celui de l'Electronique-Microélectronique (la comparaison est rendue possible par l'utilisation des indicateurs normalisés pour calculer les moyennes).

Les écarts types sont assez élevés et plus particulièrement pour les centres de recherche et les entreprises non PME. Les entreprises non PME sont toutes les entreprises de plus de 250 salariés. C'est la seule information que nous possédons sur leur taille. Cet ensemble est assez hétérogène puisqu'il regroupe aussi bien des entreprises à rayonnement régional que des filiales de grands groupes à rayonnement mondial. Il n'est donc pas étonnant que les écarts types soient élevés. Lorsque nous avons identifié les agents les plus centraux, nous avons vu que de grands groupes en faisaient partie. Si l'on avait la possibilité de différencier plusieurs catégories au sein de la catégorie IND, il serait possible que les indicateurs associés aux grands groupes soient plus élevés. La remarque est la même concernant les centres de recherche.

Lorsque l'on étudie les deux tableaux ci-dessus, on observe que les remarques que l'on peut faire sont similaires entre les deux domaines technologiques. Nous avons déjà remarqué que leurs réseaux respectifs présentaient des propriétés similaires. Il semble que dans les deux cas, le rôle de chaque type d'agent est identique. Dans les deux cas, nous observons que les agents de l'enseignement supérieur sont les agents les plus participatifs et sont ceux qui servent de relais au sein des réseaux.

En conclusion, les indicateurs de propriétés générales des réseaux ont permis de mettre en avant des propriétés favorables à l'innovation. Les réseaux de collaborations de l'UE forment des espaces de recherche intégrés dans lesquels on dénombre de nombreuses collaborations. Ces espaces se caractérisent par des distances entre les agents faibles et par le fait que le réseau constitue une composante connexe ou que la grande composante contient plus de 95% des agents. L'ensemble de ces propriétés est favorable à l'innovation. De plus, par l'intermédiaire de la comparaison entre deux domaines technologiques, nous avons pu observer que les deux réseaux possèdent des propriétés similaires. Nous avons montré que l'agrégation des deux réseaux débouche sur un réseau de taille supérieure mais avec des propriétés presque identiques aux précédents. Toutefois, nous avons pu observer que ces propriétés ne sont pas meilleures que celles que l'on aurait obtenu à partir d'un réseau aléatoire constitué des mêmes agents malgré les incitations de l'UE. Afin d'établir réellement que ces réseaux tendent vers un espace intégré donc vers un EER, il manque une dimension temporelle qui permettrait d'observer l'évolution de ces réseaux. Une véritable appréciation des PCRDT suppose, en complément de l'analyse menée, une analyse temporelle. Enfin, la portée de ces conclusions est pour l'instant limitée aux deux domaines technologiques étudiés voire par extension à la thématique IST.

L'UE a réussi à attirer au sein des réseaux de collaborations des agents qui sont des leaders dans les domaines étudiés, mais aussi des maillons essentiels du réseau comme l'ont montré les différents indicateurs de centralité mobilisés. Ces agents sont des multinationales, des centres de recherche nationaux comme le CNRS en France ou le centre EMIC en Belgique ou encore des universités (la plupart étant localisées dans les grandes capitales européennes). Parvenir à attirer au sein des réseaux de collaborations des agents leaders dans leur domaine et, en plus, les faire participer de manière active aux réseaux permet d'accroître l'attractivité des PCRDT suivants et de rendre ces réseaux plus crédibles. De plus, c'est surtout un signe d'efficacité des PCRDT puisque une partie des agents les plus compétents en font partie, cela ne pouvant être que bénéfique aux autres agents. Enfin, ces agents permettent aussi de créer des interactions entre les domaines technologiques puisqu'on les retrouve dans les deux réseaux. Il manque cependant là aussi une comparaison temporelle pour vérifier si d'un programme à l'autre les PCRDT parviennent à fidéliser les agents centraux et si possible arrivent à en attirer de nouveaux.

L'UE veut inciter les PME à participer et favoriser l'intégration des agents des nouveaux États membres. Aucune PME ne figure parmi les agents les plus centraux ce qui est compréhensible. Le bilan pour les PME est mitigé car,

si on en retrouve un grand nombre dans les deux domaines technologiques (entre 20 et 30% selon le domaine), le rôle qu'elles occupent au sein du réseau est limité du fait d'un faible nombre de participations. Toutefois, il est normal que les PME participent à moins de projet et occupent une position moins centrale car elles ne disposent pas d'autant de moyens ni d'autant de mains d'œuvre que les entreprises de plus grande taille.

Il est difficile de juger l'intégration au sein des réseaux des agents localisés dans les nouveaux États membres à partir des tableaux répertoriant les agents les plus centraux. Pour la plupart, ces agents comptent peu de participations. Il est préférable de travailler sur des données plus agrégées pour étudier de manière plus globale l'intégration de ces États à l'EER. C'est l'objet de la section 2.3.

Avant de traiter de cette question, nous apportons à l'analyse des réseaux menée jusqu'à présent une dimension temporelle afin d'observer l'évolution des programmes cadres.

2.2 Analyse temporelle des réseaux des PCRDT : application au sous-domaine des Nanotechnologies

A la section précédente, nous avons présenté et étudié les propriétés des réseaux des PCRDT de deux domaines technologiques et nous avons montré qu'elles étaient favorables à l'innovation et à la création de connaissances. Toutefois, il était impossible de dire à partir de cette analyse statique si les réseaux ont évolué vers une meilleure cohésion et si les agents se rapprochaient les uns des autres. La réponse à ces questions nécessite une analyse temporelle.

Le plus intéressant serait de pouvoir comparer les réseaux des deux domaines technologiques retenus avec leurs équivalents dans le 5^{ème} PCRDT et des programmes cadres plus anciens. Toutefois, il est difficile de dégager une telle équivalence sur l'ensemble des PCRDT du fait que les choix de thématique diffèrent à chaque programme cadre ainsi que leur décomposition en domaines technologiques. De plus, il existe aussi des problèmes d'équivalence dans les instruments utilisés et dans certaines définitions. Par exemple, dans le 4^{ème} PCRDT, un agent comptant moins de 500 employés est compté comme une PME alors que ce seuil a été abaissé à 250 à partir du 5^{ème} PCRDT.

Afin d'effectuer la comparaison la plus pertinente possible nous ne nous baserons donc que sur le 5^{ème} PCRDT (1998-2002) et le 6^{ème} PCRDT (2002-

2006). Il est difficile d'établir à partir des informations que nous possédons des équivalences entre les domaines technologiques. Les seules données pour lesquelles nous possédons une correspondance sont celles portant sur les nanotechnologies qui font partie du domaine Electronique-Microélectronique. C'est pourquoi notre analyse temporelle se focalise sur le sous-domaine des nanotechnologies que porte notre analyse temporelle.

2.2.1 Un réseau de plus en plus structuré

Les deux programmes cadres sont de dimensions différentes. Il est donc important de rappeler les écarts de budget afin de prendre en compte cet élément dans l'analyse. Les principaux éléments budgétaires (exprimés en milliards d'euros) concernant ces deux programmes cadres sont résumés dans le tableau ci-dessous.

Décomposition du budget	5 ^{ème} PCRDT	6 ^{ème} PCRDT	Évolution (en %)
Mise en œuvre des PCRDT	13,7	17,9	+ 30,5 %
Dont pour IST	3,6	3,625	+ 0,694 %
Budget total des PCRDT (y compris EURATOM)	14,96	19,113	+ 27,761 %

Les chiffres sont en milliards d'euros.

**Tableau 23 : Évolution du budget alloué aux PCRDT entre le
5^{ème} et le 6^{ème} PCRDT.**

Nous avons construit deux réseaux de collaborations correspondant aux sous domaines des nanotechnologies, un réseau pour chaque programme cadre étudié. Nous nous intéressons aux propriétés générales de ces réseaux. Il est important de vérifier si le réseau formé par les collaborations tend à se renforcer et à se consolider. Les résultats obtenus sont présentés dans le tableau suivant.

Indicateurs	FP5 Nano	FP6 Nano	Evolution en %
Nombre d'agents	379	522	+ 53%
Nombre de projets	119	74	-38%
Nombre de projets par agent	1,62	1,62	0 %
Nombre d'agents par projet	5,15	11,46	+ 123 %
Nombre de participations	613	848	+ 38 %
Nombre de liens	1619	6591	+ 307 %
Nombre de collaborations totales	1724	6977	+ 305 %
Nombre de liens potentiels	71631	135981	+ 90 %
Densité (en %)	2,26%	4,85%	+ 115 %
Intensité moyenne	1,065	1,059	- 0,1 %
Nombre de composantes	19	1	- 95 %
Taille de la grande composante	327	522	+ 60 %
en % des agents	86,3%	100%	-
en % des liens	95,8%	100%	-
Diamètre	8	5	- 37,5%
Distance moyenne	3,158	2,357	- 25 %
Coefficient de cluster	0,865	0,892	-
Centralité de degré du graphe	20,60%	43,11%	+ 105 %

Tableau : 24 Propriétés générales du réseau des Nanotechnologies pour le 5^{ème} PCRDT et le 6^{ème} PCRDT.

Dans le tableau, nous observons que le nombre de projets financés a fortement diminué alors que dans le même temps le nombre d'agents qui participent aux programmes cadres a fortement augmenté. Ceci s'explique par l'apparition de projets d'envergure : les projets NOE et IP. Ces derniers sont au cœur de la stratégie de l'UE. Ils attirent à la fois un grand nombre d'agents et une grande part du financement (74% du financement total demandé alors qu'ils représentent 39% des projets) dans le cadre d'un financement total stable. Parallèlement, nous observons que la hausse du nombre de participations ainsi que celle du nombre d'agents sont supérieures à la hausse du budget (respectivement +38% et +53%). Ceci signifie en particulier que le financement par participation a diminué entre les deux PCRDT dans IST. Ce phénomène suggère que les agents qui participent aux PCRDT cherchent de plus en plus, au-delà d'un financement, à développer des partenariats en R&D.

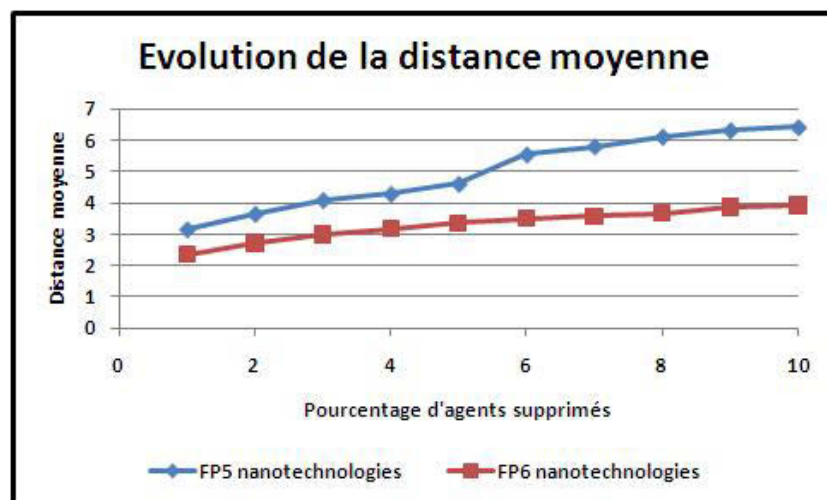
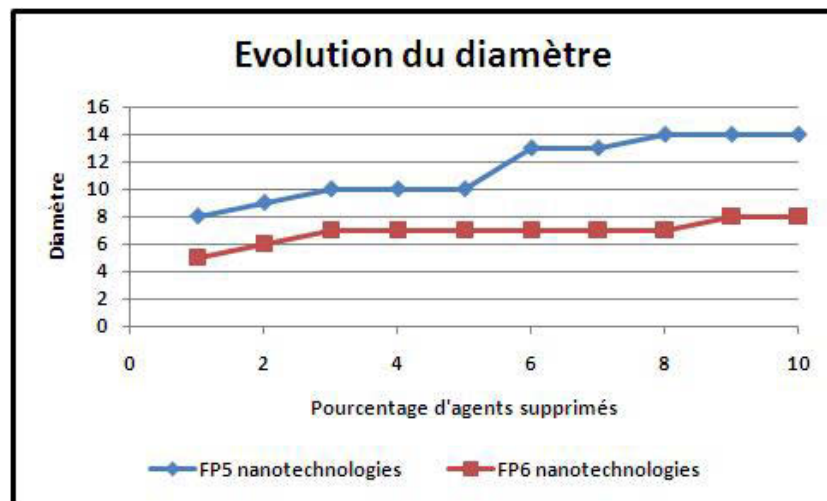
L'étude des propriétés générales du graphe semble indiquer que le réseau de collaborations tend vers un espace de plus en plus intégré et collaboratif. En premier lieu, on remarque que le nombre d'agents a fortement augmenté

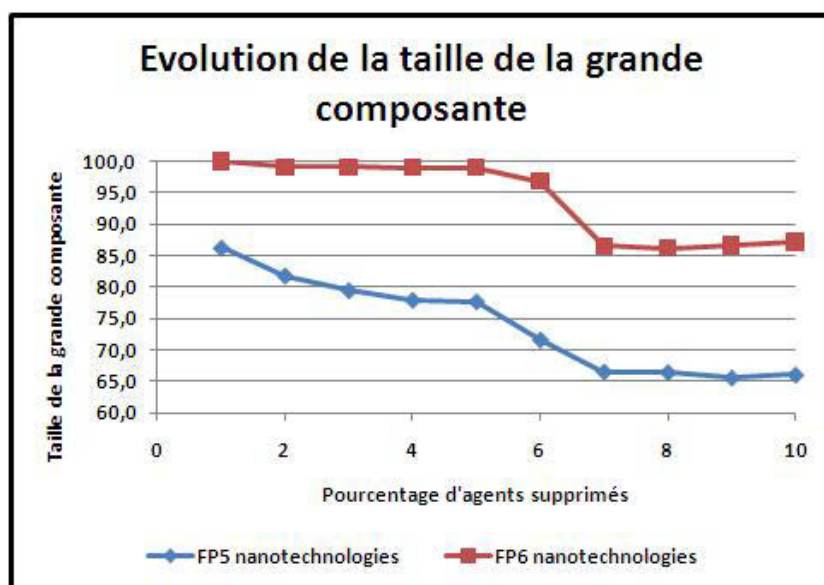
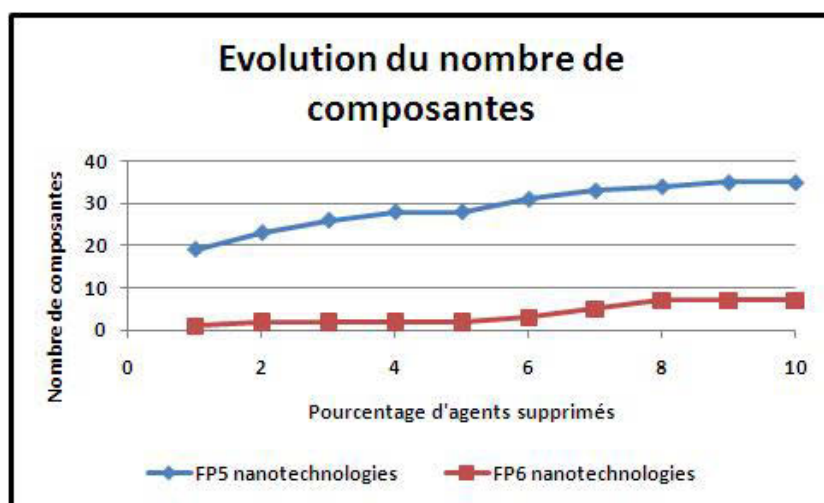
entre les deux programmes cadres ce qui répond à l'objectif d'attirer plus de partenaires dans le réseau. En second lieu, on observe une hausse du nombre d'agents au sein de chaque projet. La taille des projets a plus que doublé en moyenne entre les deux programmes cadres. Ceci traduit une hausse du nombre de collaborations entre les agents qui constitue un des objectifs fixés aux programmes cadres. Cette augmentation de la taille des projets est une manière d'atteindre pour chaque projet une masse critique suffisante pour mener une recherche d'envergure notamment dans les projets IP et NOE. Enfin, la connectivité entre les agents mesurée par les indicateurs de densité et de distance moyenne est plus forte. Concernant la densité du graphe, cette dernière s'est accrue, les agents réalisent 4,85% des liens possibles dans le 6^{ème} PCRDT contre seulement 2,26% dans le 5^{ème} PCRDT. Ceci s'explique par une hausse du nombre de participations, une hausse du nombre d'agents et, en parallèle, la baisse du nombre de projets financés. En outre, les agents sont plus proches les uns des autres et pourraient profiter plus facilement des externalités de connaissance produites par les autres projets. En effet, la distance moyenne entre les sommets du graphe a diminué passant de 3,158 à 2,357, baisse qui s'explique en partie par l'augmentation du nombre d'agents par projet ce qui induit un plus grand nombre de paires d'agents à distance 1. En parallèle, le diamètre du graphe est passé de 8 à 5 ce qui signifie que le chemin le plus long entre deux sommets du graphe est de 5 (ce qui tend à montrer que la baisse de la distance moyenne entre agents n'est pas due qu'à une hausse du nombre d'agents par projet). Cette notion de distance est d'autant plus importante si on suppose que les externalités de connaissance s'atténuent avec la distance (hypothèse de decay). Enfin, le réseau formé par le 6^{ème} PCRDT ne constitue plus qu'une seule composante contre 19 auparavant. Tous les agents peuvent donc communiquer entre eux, au moins de manière indirecte, facilitant les collaborations et la transmission de connaissances.

2.2.2 Un réseau plus robuste

A la sous-section précédente, nous avons observé que les propriétés générales du réseau nanotechnologies ont évolué dans le sens souhaité par l'UE, c'est-à-dire que le réseau tend à devenir un espace intégré avec une seule composante de plus en plus collaborative et dans lequel les distances entre les agents sont plus faibles. Dans le tableau précédent, nous avons aussi observé que la centralité de degré du graphe a augmenté. Ceci signifie que des agents jouent un rôle de "hubs" au sein du réseau ce qui explique en partie le gain d'efficacité du réseau et la meilleure connexion de celui-ci. Si toutes

ces propriétés sont favorables à l'innovation, il reste néanmoins à vérifier que l'émergence de "hubs" (les agents qui possèdent le plus de liens) ne fragilise pas le réseau en le rendant plus dépendant de ces derniers. Pour vérifier cela, nous procédons à la suppression séquentielle par tranche de 1% des "hubs" de chaque réseau afin de tester la robustesse de ces derniers. Les résultats sont présentés en annexes A.7. L'évolution des principaux indicateurs est visible sur les graphiques ci-dessous.





Graphiques 29 à 32 : Évolution des propriétés générales des réseaux nanotechnologies suite à la suppression des agents les plus centraux.

Quel que soit l'indicateur, le réseau des nanotechnologies dans le 6^{ème} PCRDT apparaît plus robuste. En effet, la distance moyenne et le diamètre déjà inférieurs augmentent certes mais moins que dans le réseau des nanotechnologies du 5^{ème} PCRDT. Ceci est valable aussi bien dans l'absolu que de manière relative. Le diamètre du réseau pour le 5^{ème} PCRDT passe de 8 à 14 soit une augmentation de 75% (+6) alors que celui du 6^{ème} PCRDT augmente de 60% (+3). Le résultat est identique pour la distance moyenne

qui fait plus que doubler dans le réseau du 5^{ème} PCRDT alors qu'elle n'augmente que de 67% dans le 6^{ème} PCRDT. Le réseau du 5^{ème} PCRDT éclate au fur et à mesure de la suppression des agents les plus centraux pour finalement compter 35 composantes après suppression de 10% des "hubs" contre seulement 7 pour le réseau du 6^{ème} PCRDT. De plus, la plus grande composante du 5^{ème} PCRDT ne compte plus que 66% des agents restants contre 88% dans le réseau du 6^{ème} PCRDT. On peut remarquer que pour les deux réseaux la suppression de 7% des agents constitue un seuil puisque dans les deux cas le pourcentage supplémentaire se traduit par une forte baisse de la taille de la plus grande composante.

La limite de notre analyse tient à la taille relativement faible des deux réseaux et au nombre réduit d'agents. Ceci explique en partie que ces réseaux soient plus sensibles que les réseaux des deux domaines technologiques étudiés précédemment. De plus, le fait que ces deux réseaux soient construits sur un faible nombre de projets introduit un biais important plus particulièrement pour celui du 6^{ème} PCRDT. En effet, ce réseau compte des projets d'envergure composés de nombreux agents. Ces projets sont générateurs d'énormément de liens. Un agent qui ne participe qu'à un ou deux projets peut alors être considéré comme "hubs", c'est-à-dire un agent doté de nombreux liens si le peu de projets auquel il participe sont des projets NOE ou IP. L'indicateur de degré est plus sujet à des biais lorsque le nombre de projets et le nombre d'agents sont faibles. Par exemple, on peut remarquer dans l'annexe A.7 que pour le réseau du 6^{ème} PCRDT, le passage de 6% à 7% d'agents supprimés impacte entièrement la seconde plus grande composante et non la plus grande composante. Ceci s'explique par le fait que parmi les "hubs" figurent des agents qui ne participent qu'à un seul projet mais un projet de grande taille. Malgré cela étant donné la définition courante d'un "hub" ces agents restent des agents centraux. Ce biais rend donc l'analyse moins pertinente. Le biais est beaucoup moins important pour les deux domaines technologiques qui comptent un plus grand nombre de projets et d'agents et pour lesquels nous avons vérifié au préalable qu'aucun des "hubs" n'étaient en fait un agent mono projet.

Nous étudions maintenant la continuité entre les deux réseaux.

2.2.3 Des agents de plus en plus impliqués

La seule possibilité pour étudier la continuité entre les deux réseaux est de s'intéresser aux agents qui participent à la fois au réseau des nanotechnologies dans le 5^{ème} PCRDT et dans le 6^{ème} PCRDT. Les projets ne peuvent en aucun cas être commun à deux programmes cadres. Sur les 379 agents qui

composent le réseau des collaborations dans le domaine des nanotechnologies du 5^{ème} PCRDT, 116 participent aussi à celui du 6^{ème} PCRDT soit 31%. Ces agents représentent 43% des participations du 5^{ème} PCRDT (262 sur 613) et 42% de celles du 6^{ème} PCRDT (357 sur 848). La participation de ces agents aux deux réseaux est donc stable. Ces 116 agents se décomposent de la manière suivante :

Type 1	Variation des participations			Total	En % du total d'agents
	Egal	En progrès	En retrait		
HE	16	20	6	42	61%
IND	18	10	7	35	21%
IND PME	10	1	1	12	55%
OTH	3	0	2	5	8%
RES	5	11	6	22	35%
Total général	52	42	22	116	31%

Tableau 25 : Répartition des agents participants aux deux programmes cadres dans le domaine des nanotechnologies selon leur type et la variation du nombre de leurs participations.

Les agents de l'enseignement supérieur et les PME sont les plus nombreux à renouveler leur participation d'un programme à l'autre. Au contraire, peu d'entreprises non PME renouvellent leur participation. Il faut noter que puisque nous raisonnons à partir des établissements/filiales et non des groupes (chaque établissement/filiale d'un groupe est un agent différent), il est possible que d'un programme cadre à l'autre le groupe continue de participer au sous-domaine des nanotechnologies. Les agents de la catégorie autres sont peu nombreux à participer aux deux programmes cadres dans le domaine des nanotechnologies. Les agents de l'enseignement supérieur sont, avec les centres de recherche, ceux qui ont le plus accru leur participation dans le domaine des nanotechnologies.

S'il est important d'attirer de nombreux agents, il est aussi important de conserver d'un programme à l'autre les agents les plus centraux. Nous avons vu, à la sous-section précédente, que le faible nombre de projet du réseau nanotechnologies du 6^{ème} PCRDT pouvait introduire un biais dans l'identification des "hubs". Dans notre cas, nous avons défini un agent central comme un agent devant répondre à 5 critères. Le fait d'avoir utiliser ces 5 critères permet de limiter ce biais voire de l'éliminer totalement. Étant donné la taille plus réduite des réseaux, nous ne présentons que les 20 agents les plus centraux (cf. annexes A.8.).

Si l'on s'intéresse aux agents les plus centraux on observe que d'un programme à l'autre on retrouve des agents similaires parmi les agents les plus centraux. Au total, on en dénombre 7 parmi les 20 agents les plus centraux et, en élargissant, 27 agents en commun parmi les 100 agents les plus centraux. Ce chiffre peut paraître assez faible car on aurait pu penser que d'un programme à l'autre les agents centraux soient quasiment identiques.

D'une certaine façon, il est possible de dire que ce chiffre est sous-estimé. En effet, même si on s'intéresse aux collaborations entre agents et que l'on définit un agent comme un établissement et non un groupe, nous sommes obligés de constater que de nombreux agents parmi les plus centraux font partie d'ensemble plus large. Parmi ces ensembles figurent les grands groupes industriels mondiaux tels que Philips, Alcatel mais aussi des laboratoires de recherche assez importants tels que Fraunhofer Gesellschaft zur Forderung, IMEC, le CEA, le CNRS, ... Tous ces groupes ont la particularité de posséder un agent parmi les agents les plus centraux de chaque réseau mais cet agent est différent d'un programme à l'autre. Si l'on prend l'exemple du CEA (Commissariat à l'Énergie Atomique), c'est la structure localisée à Grenoble qui participe aux 5^{ème} PCRDT dans le domaine des nanotechnologies alors qu'il s'agit de celle localisée à Paris qui figure parmi les agents les plus centraux du 6^{ème} PCRDT. Plusieurs raisons pourraient expliquer ce phénomène : satisfaire les conditions imposées par l'UE pour chaque programme en termes de nationalité des agents, des raisons liées à la proximité géographique, le fait que le projet dans lequel travaille une filiale n'est pas encore achevé ou encore des spécialités différentes d'une filiale à l'autre qui correspondent mieux à certains projets que d'autres. Il serait intéressant ainsi de travailler à un niveau plus agrégé, c'est à dire au niveau des groupes afin de voir si ces derniers prennent plus d'ampleur au sein du réseau. Il faut toutefois retenir qu'il est possible qu'il existe au niveau des PCRDT des stratégies de groupe qui dépassent les stratégies individuelles de collaborations.

Il existe des difficultés pratiques pour travailler sur les groupes, en particulier des problèmes d'identification. Pour identifier les groupes, la seule information dont nous disposons c'est le nom. En plus des erreurs d'orthographe, des fautes de frappe, ... qui empêchent une identification précise des agents, le nom des agents présents dans la base de données peut parfois être complètement différents de celui du groupe ce qui renforce les problèmes d'identification.

Enfin, nous observons que, dans le classement des agents les plus centraux du 5^{ème} PCRDT, ne figure aucune PME ni aucun agent localisé dans l'un des nouveaux Etats membres. Par contre, dans le classement du 6^{ème} PCRDT, s'il n'y a toujours aucune PME, on note l'apparition de deux agents localisés dans des nouveaux Etats membres. Il s'agit de l'institut de technologie

électronique et de l'université technique de Varsovie tous les deux localisés en Pologne.

Les PCRDT parviennent donc à fidéliser d'un programme à l'autre un certain nombre d'agents. Il est aussi important de montrer qu'ils attirent de nouveaux partenaires. Ainsi même s'ils ne figurent pas dans le classement des agents les plus centraux, il faut noter par exemple qu'entre le 5^{ème} PCRDT et le 6^{ème} PCRDT, le groupe EADS fait partie des nouvelles entreprises qui ont intégré les PCRDT. De plus, certains agents déjà présents au 6^{ème} PCRDT, accroissent leur participation de manière significative dans le 6^{ème} PCRDT comme l'université Catholique de Louvain qui fait partie des agents les plus centraux. Enfin, certains acteurs principaux du réseau (qui faisaient partie des 20 agents les plus centraux du 5^{ème} PCRDT) ont accru leur participation dans le 6^{ème} PCRDT comme Alcatel par exemple sans pour autant figurer parmi les plus centraux du 6^{ème} PCRDT. Les PCRDT arrivent donc à conserver d'un PCRDT à l'autre un ensemble d'agents et/ou de groupes qui constituent le cœur du réseau. Ceci permet de pouvoir attirer d'un programme à l'autre des partenaires susceptibles d'envisager des collaborations avec les agents les plus importants du précédent programme cadre. Enfin, cette fidélité d'un programme à l'autre est aussi la preuve d'une certaine efficacité des PCRDT à assurer leur mission première, c'est-à-dire la production de savoirs et l'innovation. Les plus grands groupes industriels continuent à participer, signe qu'il est intéressant pour eux de figurer et d'intégrer ces programmes cadres. Ces agents ont pour la plupart des objectifs de rentabilité, on peut donc supposer que s'ils participent aux PCRDT c'est que par la suite ils arrivent à exploiter les travaux effectués dans le cadre des PCRDT.

2.2.4 Des PME intégrées au réseau des nanotechnologies

A partir de la comparaison entre le réseau des nanotechnologies du 5^{ème} PCRDT et du 6^{ème} PCRDT, il est aussi possible de s'intéresser à l'évolution de la place occupée par les PME au sein du réseau. Nous avons vu à la section précédente que ces agents participaient souvent peu aux programmes cadres (une ou deux participations). Nous cherchons alors à évaluer de quelle manière leur position a évolué au sein du réseau entre les deux programmes cadres à l'aide des indicateurs de centralité (même si nos conclusions ne seront valables que dans le cas des nanotechnologies). Pour cela, nous calculons pour chaque type d'agent un indicateur moyen pour le 5^{ème} PCRDT et le 6^{ème} PCRDT. De manière à pouvoir réaliser cette comparaison, les indicateurs

moyens sont calculés à partir des mesures normalisées. Les résultats sont présentés dans les tableaux ci-dessous.

FP5 Nanotechnologies						
Type 1	Nb d'ag.*	Proj. Norm.**	CD	CC	CB	CE
HE (EC***)	64	1,43 (1,15)	3,69 (2,89)	34,22 (5,14)	0,59 (1,29)	5,74 (6,43)
IND (EC***)	143	1,32 (1,32)	2,53 (2,73)	31,82 (5,28)	0,34 (1,49)	3,34 (6,79)
IND PME (EC***)	17	1,24 (0,60)	2,18 (1,46)	30,70 (5,31)	0,34 (0,72)	2,07 (2,84)
OTH (EC***)	49	0,96 (0,30)	1,77 (1,12)	31,02 (5,12)	0,08 (0,45)	2,24 (2,96)
RES (EC***)	54	2,27 (3,03)	4,26 (4,45)	34,97 (5,19)	1,21 (3,18)	6,02 (9,02)
MG**** (EC***)	327	1,44 (1,64)	2,91 (3,01)	32,63 (5,40)	0,49 (1,76)	4,02 (6,72)

(* Nb d'ag. = Nombre d'agents; ** Proj. Norm. = Mesure du nombre de projets normalisés; *** EC = Ecart-type; **** MG = Moyenne générale)

Type 1	FP6 Nanotechnologies				
	Nb d'ag.*	Proj. Norm.**	CD	CC	CB
HE (EC***)	133	2,47 (2,04)	5,99 (4,81)	44,32 (5,13)	0,34 (0,83)
IND (EC***)	118	2,22 (2,56)	4,47 (3,97)	43,51 (4,67)	0,14 (0,57)
IND PME (EC***)	140	1,48 (0,58)	3,08 (2,52)	41,24 (3,85)	0,02 (0,13)
OTH (EC***)	37	1,61 (0,62)	4,06 (2,85)	42,03 (2,81)	0,01 (0,04)
RES (EC***)	94	3,08 (5,57)	6,64 (7,96)	43,55 (6,41)	0,75 (3,01)
MG**** (EC***)	522	2,20 (2,92)	4,85 (4,98)	43,01 (4,99)	0,26 (1,39)
					CE
					4,53 (4,31)
					4,11 (4,92)
					2,21 (2,12)
					2,58 (2,01)
					5,11 (7,93)
					3,78 (4,91)

(* Nb d'ag. = Nombre d'agents; ** Proj. Norm. = Mesure du nombre de projets normalisés; *** EC = Ecart-type; **** MG = Moyenne générale)

Tableaux 26 et 27 : Indicateurs de centralité moyen pour chaque type d'agents.

La première remarque est que le nombre de PME a fortement augmenté dans le sous-domaine des nanotechnologies entre les deux programmes cadres. Alors que l'on comptait seulement 17 PME dans le 5^{ème} PCRDT, on en dénombre 140 dans le 6^{ème} PCRDT.

Tous les indicateurs de centralité retenus (sauf celui de betweenness) ont augmenté ce qui signifie que les PME sont plus centrales dans le réseau nanotechnologies du 6^{ème} PCRDT que dans celui du 5^{ème} PCRDT. En observant plus attentivement ces indicateurs, on s'aperçoit que le constat est le même pour tous les autres type d'agents. Les indicateurs des PME suivent une évolution semblable à celle des autres agents du réseau.

En moyenne, une PME participait à 1,24% des projets dans le 5^{ème} PCRDT contre 1,48% dans le 6^{ème} PCRDT. Le nombre de projets a fortement diminué entre les deux programmes cadres passant de 119 à 74. Au final, le nombre moyen de projets auxquels participent les PME a donc diminué. Le nombre de partenaires moyens n'a fait que suivre l'évolution du nombre d'agents ainsi que l'accroissement du nombre d'agents par projets créant de ce fait un nombre de liens plus important.

Le seul indicateur qui n'a pas augmenté entre les deux programmes cadres pour les PME est celui de betweenness (ou centralité d'intermédiation). Alors qu'il était identique à celui des entreprises non PME dans le 5^{ème} PCRDT, il est inférieur dans le 6^{ème} PCRDT. En termes de centralité d'intermédiation, les PME sont aussi centrales que les agents de la catégorie autres. Cette diminution de la centralité d'intermédiation signifie donc que les PME n'occupent plus une position de relais aussi importante dans le réseau du 6^{ème} PCRDT que dans le 5^{ème} PCRDT.

On observe que les indicateurs de betweenness et de centralité par les valeurs propres sont aussi en baisse pour les autres types d'agents (la baisse est toutefois plus visible pour les PME). Les nouveaux instruments (NOE et IP) en créant d'importants consortia (se traduisant par des cliques sur le graphe) ont réduit l'importance des principaux agents. Le nombre d'agents par projet étant plus élevé et le nombre de projets plus faible, cela se traduit par une hausse du nombre d'agents en commun entre deux projets. De ce fait, il existe plus d'un intermédiaire entre chaque projet ce qui réduit l'indicateur de betweenness de chaque agent.

Cette hausse du nombre de PME entre les deux programmes cadres peut être vue comme le souhait de l'UE de faire participer de manière plus intensive ces agents aux programmes cadres. Pourtant les PME, bien que moins nombreuses dans le 5^{ème} PCRDT, semblaient mieux intégrées. Les dispositifs incitatifs mis en place par l'UE pour favoriser leur participation ont peut être incité les différents consortiums à intégrer plus de PME au sein des projets

afin de pouvoir faciliter le financement du projet. Une étude économétrique portant sur le financement ou non des projets en fonction de critères scientifiques mais aussi de critères liés à la participation des PME pourraient permettre à l'avenir de vérifier l'impact de la participation de ces dernières.

Au final, il ressort de cette comparaison temporelle trois principaux résultats.

Premier résultat, le réseau des nanotechnologies tend à se consolider. Ce sous-domaine attire de plus en plus d'agents. Les distances entre les agents tendent à diminuer et les collaborations à augmenter. De plus, le réseau tend à devenir moins indépendant des agents les plus centraux. Toutefois, une partie de ces résultats semblent être due à la création de projets d'envergure type NOE ou IP. Ces projets par leur taille génèrent un grand nombre de liens et tendent donc à réduire les distances entre agents. Il est difficile d'établir la part des résultats relative à ces projets.

Deuxième résultat, le réseau des nanotechnologies conserve une grande partie de ses agents d'un programme à l'autre. De plus, il conserve aussi une grande partie de ses agents les plus centraux. Enfin, si certains agents cessent leur participation d'un programme à l'autre ou diminuent leur implication, nous avons aussi mis en avant des stratégies de groupe. Dans ces cas, nous avons observé qu'une partie des agents était remplacée au programme cadre suivant par des agents du même groupe mais localisés à des endroits différents.

Troisième résultat, le nombre de PME participant aux programmes cadres s'est accru dans le réseau des nanotechnologies du 6^{ème} PCRDT. Nous avons vu que les PME sont le type d'agent dont les indicateurs de centralité de *betweenness* et d'*eigenvector centrality* ont le plus fléchi entre les deux programmes cadres ce qui nous laisse supposer que l'influence des PME au sein du réseau a diminué.

Rappelons que ces résultats ne sont valables que pour le réseau des nanotechnologies et demanderaient à être vérifiés à un niveau plus global.

Jusqu'à présent, nous nous sommes intéressés aux agents de manière individuelle. Les indicateurs de centralité utilisés dans notre analyse sont conçus avant tout pour étudier la position occupée par les agents dans un réseau. Toutefois, il peut apparaître opportun dans le cas des PCRDT d'analyser des réseaux constitués à un niveau plus agrégé. Ainsi, la plupart des indicateurs déployés jusque là ne nous ont pas permis d'apprécier la réalisation d'un des objectifs conduits par l'UE au sein de sa politique des PCRDT à savoir l'intégration d'agents des nouveaux États membres dans le 6^{ème} PCRDT et un développement harmonieux sur l'ensemble du territoire européen. Nous avons remarqué que certains de ces agents étaient parmi les plus centraux

sans pouvoir jusqu'à présent déterminer s'il s'agissait de quelques agents isolés ou au contraire d'une réelle intégration au sein des réseaux. Afin de fournir des éléments d'appréciation sur ces objectifs, nous montrons dans la section suivante comment l'analyse réseau, au travers des indicateurs déployés dans le cas de l'étude des collaborations individuelles, peut permettre de construire et d'analyser des réseaux de régions européennes.

2.3 Une agrégation des données au niveau régional pour compléter l'évaluation.

Dans les deux sections précédentes, nous avons pu remarquer la présence parmi les agents les plus centraux de quelques agents localisés dans les nouveaux États membres. Toutefois, il est difficile, à partir de ces cas isolés, de déterminer réellement de quelle manière les nouveaux États membres sont intégrés aux PCRDT.

Les réseaux de collaborations individuelles dans le cas d'une politique comme les PCRDT ne permettent pas de vérifier l'ensemble des objectifs fixés. C'est notamment le cas de ceux portant sur l'implication des différentes régions européennes et plus particulièrement celles issues des nouveaux États membres de l'UE. Cet objectif est clairement affiché par l'UE au moment du lancement du 6^{ème} PCRDT dans un document intitulé "Integrating and strengthening the European Research Area" (2004, p4) : *"this work program underlines the importance of involving associated candidate countries in the Community's research policy and in the European Research Area. Specific support actions will also be implemented to stimulate, encourage and facilitate the participation of organisations from the candidate countries in the activities of the priority thematic areas"* (Source Cordis). L'EER, tel qu'il est souhaité par l'UE, se veut un espace de collaborations entre agents mais aussi un espace où un grand nombre de régions européennes sont impliquées, en particulier les régions des nouveaux États membres. Les PCRDT doivent favoriser les collaborations entre des agents provenant de régions éloignées et n'ayant pas forcément pour habitude de collaborer ensemble. Pour apprécier l'implication des différentes régions européennes, il est nécessaire de construire des réseaux plus agrégés et de les étudier de manière similaire. Les NUTS⁶ permettent de donner une dimension géographique à l'étude en s'intéressant aux régions impliquées dans cet EER et en dégagant les axes principaux de collaboration au sein de l'Europe. Cette information n'apparaît pas dans les réseaux de collaborations entre agents.

⁶NUTS : Nomenclature d'Unités Territoriales Statistiques.

Dans cette section, nous nous intéressons à la participation des nouveaux États membres au sein des deux domaines technologiques étudiés à la section 2.1. Dans un premier temps, nous présentons quelques indicateurs de participations de ces pays. Dans un second temps, nous expliquerons pourquoi il nous semble plus pertinent de travailler sur les régions européennes plutôt qu’au niveau des pays afin d’étudier l’intégration des nouveaux États membres aux PCRDT. En cherchant à vérifier la mise en œuvre des objectifs de l’UE, nous montrerons l’adaptabilité de l’analyse réseau dans l’étude de réseaux de collaborations diverses.

2.3.1 Les nouveaux États membres, des pays en retrait

Afin d’étudier la place des nouveaux États membres, nous avons calculé pour chaque pays trois indicateurs différents :

- le nombre d’agents ;
- le nombre de participations réalisées par des agents locaux ;
- un indicateur de participation moyenne qui est égal au rapport entre le nombre de participation du pays et le nombre d’agents du pays.

Ces trois indicateurs ont été calculés pour chaque pays et pour les trois réseaux étudiés, c’est-à-dire le réseau Electronique-Microélectronique, Télécommunications et le réseau agrégé. Les résultats sont présentés dans l’annexe A.9.

Les pays de l’Union Européenne à 15 membres (UE15) concentrent la plupart des agents qui participent aux programmes cadres mais aussi la plupart des participations. En effet, les pays de l’UE des 15 représentent plus de 80% des agents quel que soit le domaine technologique étudié et réalisent plus de 85% des participations. En comparaison, la part des agents issus des Nouveaux États Membres (NEM) est d’environ 7% et ne représente que 5,5% des participations. Le taux de participation moyen des nouveaux États membres égal à 1,89 est plus faible que celui des pays de l’UE 15 qui est de 2,68. De manière générale, on observe que les nouveaux États membres ne sont pas bien intégrés aux PCRDT.

Toutefois, il existe des disparités entre les nouveaux États membres puisque deux d’entre eux, la Pologne et la Hongrie, possèdent des indicateurs relativement proches de certains pays de l’UE15 que sont le Danemark, l’Irlande et le Portugal même s’il s’agit des pays de l’UE15 les moins participatifs. De plus, ces deux pays possèdent un taux de participation moyen relativement élevé puisqu’il est de 2,53 pour la Pologne et de 2,46 pour la Hongrie se rapprochant ainsi de celui des pays de l’UE15. Dans notre étude, ce sont donc la Pologne et la Hongrie qui ressortent. Toutefois, ceci peut provenir des

domaines technologiques étudiés alors que, dans d'autres domaines, d'autres pays pourraient émerger.

Pour compléter ces observations, nous pouvons formuler deux remarques.

La première c'est qu'un cœur⁷ composé de l'Allemagne, l'Espagne, la France, l'Italie et le Royaume-Uni semble se dégager (dans le cas du réseau Electronique-Microélectronique, la Belgique fait aussi partie de ce cœur). En effet, ces pays concentrent une grande partie des agents mais aussi des participations. Cependant, il apparaît qu'entre eux, il existe de grandes disparités. En effet, une grande partie des agents et des participations de ces pays est concentrée dans un faible nombre de régions. L'exemple le plus flagrant de ce phénomène est la France où la région Ile de France concentre à elle seule 51% des agents français et 68% des collaborations. On suppose que dans le cas des nouveaux États membres c'est aussi le cas, à savoir que le peu de participations qu'ils ont sont concentrées dans un faible nombre de régions. En étudiant les relations de collaboration au niveau des régions, il sera possible ainsi de comparer les régions des nouveaux États membres non pas avec les régions comme l'Ile de France ou la Communauté de Madrid par exemple mais avec les régions moins participatives de ces pays. L'objectif sera alors de déterminer si certaines régions des nouveaux États membres arrivent à atteindre un niveau de participation proche des régions secondaires des pays du cœur. De plus, une comparaison au niveau régional peut permettre de prendre en compte un premier effet taille. En effet, la moitié des nouveaux États membres (Malte, Chypre, Lituanie, Lettonie, Estonie et Slovaquie) est de taille relativement faible. Dans ces cas, le pays ne forme qu'un seul NUTS2. Les NUTS2 sont en fait des unités territoriales qui correspondent aux régions françaises.

La seconde remarque c'est le fait qu'il semble exister des phénomènes de spécialisation. En effet, certains pays concentrent une grande partie de leurs agents sur un domaine particulier. C'est le cas des pays qui comptent peu de participations par rapport au cœur identifié précédemment. L'exemple le plus marquant est celui du Portugal. Ce pays compte seulement 9 agents et 14 participations dans le réseau Electronique-Microélectronique alors qu'il compte 23 agents dans le réseau Télécommunications mais surtout 81 participations. En revanche, il est impossible de déterminer si ces phénomènes s'expliquent par une spécialisation de la région ou s'il s'agit d'un phénomène de rattrapage. Pour déterminer quel effet joue, il faudrait pouvoir croiser

⁷La méthode cœur/périphérie n'est pas mobilisée pour déterminer ce cœur. Il est déterminé par le nombre d'agents et de participations bien au dessus de la moyenne de ces pays.

avec d'autres bases de données du type Diane en France afin de connaître la localisation des établissements principaux français et leurs activités. Ces phénomènes de spécialisation peuvent être détectés à un niveau plus fin qui est celui des régions. L'agrégation au niveau des pays peut cacher ces phénomènes de spécialisation.

Afin d'étudier la place occupée par les nouveaux États membres, il apparaît qu'il est plus juste de réaliser l'étude à un niveau plus fin. Utiliser le niveau régional est une solution intéressante qui permet de prendre en compte à la fois un effet taille puisque tous les pays ne sont pas de taille équivalente tout en évitant que les régions les plus participatives faussent la comparaison à un niveau plus agrégé. Dans un premier temps, ils nous faut donc construire les réseaux de relations entre les régions. C'est l'objet de la section suivante.

2.3.2 Des réseaux de collaborations individuels aux réseaux de relations entre régions

Dans cette sous-section, nous montrons comment nous utilisons les réseaux de collaborations entre agents individuels afin de construire le réseau de relations entre régions. La première étape consiste à construire le réseau de collaborations entre agents dans un domaine particulier. Pour chaque agent, nous possédons l'information relative à sa localisation géographique. Dans une seconde étape, grâce à cette information, nous construisons le réseau de NUTS. Par exemple, si un agent i localisé en Ile de France collabore au sein d'un projet avec un agent j localisé dans la région Rhône-Alpes alors il existe un lien entre la région Rhône-Alpes et l'Ile de France. Dans la suite de cette section, nous décrivons de manière plus formelle la méthode qui permet de construire ces réseaux.

La première étape consiste à créer le réseau de collaborations entre agents (cf. section 1.2). Le réseau est alors construit de la manière suivante : deux agents qui participent à un même projet ont un lien entre eux. Dans notre cas, nous utiliserons une matrice d'adjacence valuée X^V (de taille $N \times N$ avec N le nombre d'agents) définie de la façon suivante :

$$x_{ij}^V = \begin{cases} k & \text{avec } k, \text{ le nombre de projets en commun entre } i \text{ et } j \text{ si } i \neq j \\ 0 & \text{si } i = j \end{cases}$$

Afin d'illustrer nos propos, prenons l'exemple suivant :

Agent	Région	Projet
1	<i>a</i>	<i>P1</i>
2	<i>b</i>	<i>P1</i>
3	<i>c</i>	<i>P1</i>
4	<i>c</i>	<i>P1</i>
5	<i>e</i>	<i>P2</i>
1	<i>a</i>	<i>P2</i>
2	<i>b</i>	<i>P2</i>
1	<i>a</i>	<i>P3</i>
2	<i>b</i>	<i>P3</i>
6	<i>a</i>	<i>P3</i>
7	<i>c</i>	<i>P4</i>
8	<i>c</i>	<i>P4</i>
9	<i>d</i>	<i>P4</i>
5	<i>e</i>	<i>P5</i>
9	<i>d</i>	<i>P5</i>

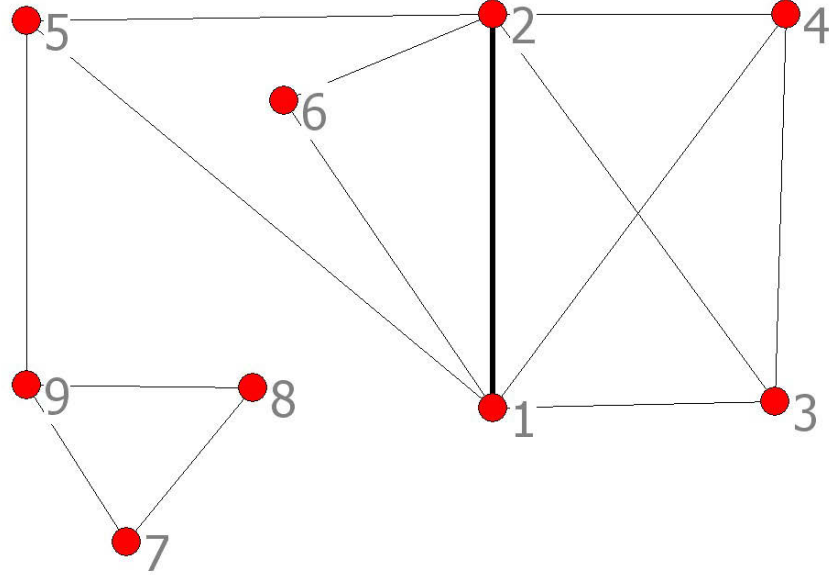
Exemple 2

La première ligne se lit de la manière suivante : l'agent 1 localisé dans la région *a* participe au projet *P1*. A partir de ce tableau, on construit la matrice de collaboration entre agents de la façon suivante.

	1	6	2	3	4	7	8	9	5
1	0	1	3	1	1	0	0	0	1
6	1	0	1	0	0	0	0	0	0
2	3	1	0	1	1	0	0	0	1
3	1	0	1	0	1	0	0	0	0
4	1	0	1	1	0	0	0	0	0
7	0	0	0	0	0	0	1	1	0
8	0	0	0	0	0	1	0	1	0
9	0	0	0	0	0	1	1	0	1
5	0	0	0	0	0	0	0	1	0

Matrice : Matrice évaluée de l'exemple décrivant les collaborations entre les différents agents.

La première ligne de la matrice se lit de la façon suivante : l'agent 1 participe à un projet avec l'agent 6, à 3 projets avec l'agent 2,... A partir cette matrice, on obtient le réseau de relations entre agents suivant :



Graphe 15 : Graphe valué de collaborations entre agents.

Ensuite, à partir des collaborations entre agents, nous construisons la matrice des relations entre région X^R (matrice carrée $R \times R$ avec R le nombre de régions). A chaque région correspond un groupe d'agents et chaque agent appartient à une région de telle manière que les intersections entre les groupes d'agents (entre les régions) soient vides. Les relations entre régions étant construites sur la base des relations entre agents, il est alors possible qu'il existe des liens entre agents d'une même région. Dans ce cas, cela crée des liens réflexifs, c'est-à-dire des liens dont le point de départ et le point d'arrivée sont la même région. Les éléments de la diagonale de la matrice X^R peuvent donc être non nuls. Les valeurs de la diagonale correspondent alors à des collaborations intra-régions.

Pour déterminer la valeur des éléments de la matrice X^R , on procède de la manière suivante. Soit x_{ab}^R correspondant à la relation entre la région R_a et la région R_b composée respectivement des groupes d'agents N_a et N_b avec

$N_a \subset N$, $N_b \subset N$ et $N_a \cap N_b = \emptyset$ pour $a \neq b$. On définit alors x_{ab}^R :

- si $a = b$ (cas des relations intra-régions) alors

$$x_{ab}^R = x_{aa}^R = \frac{\sum_{k \in N_a} \sum_{k' \in N_a} x_{kk'}^V}{2}$$

avec $k \in N_a$ et $k' \in N_a$. Dans ce cas, on a $N_a = N_b$ ⁸ ;

- si $a \neq b$ alors

$$x_{ab}^R = x_{ba}^R = \sum_{k \in N_a} \sum_{k' \in N_b} x_{kk'}^V = \sum_{k \in N_a} \sum_{k' \in N_b} x_{k'k}^V$$

avec $k \in N_a$ et $k' \in N_b$;

Dans le cas où $a \neq b$, la valeur de x_{ab}^R est égale à celle de x_{ba}^R puisque l'on considère que la relation est symétrique. En effet, dans notre cas, le lien qui va de la région a à la région b est strictement équivalent au lien qui va de la région b vers la région a .

Si l'on reprend l'exemple précédent, on obtient :

	Région a		Région b		Région c				Région d	Région e
	1	6	2		3	4	7	8	9	5
1	0	1	3		1	1	0	0	0	1
6	1	0	1		0	0	0	0	0	0
2	3	1	0		1	1	0	0	0	1
3	1	0	1		0	1	0	0	0	0
4	1	0	1		1	0	0	0	0	0
7	0	0	0		0	0	0	1	1	0
8	0	0	0		0	0	1	0	1	0
9	0	0	0		0	0	1	1	0	1
5	0	0	0		0	0	0	0	1	0

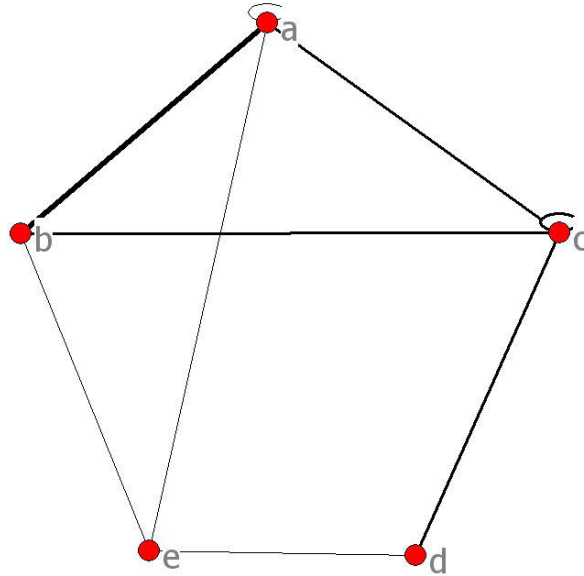
Matrice : Matrice valuée de l'exemple décrivant les collaborations entre les différents agents regroupés par région.

A partir de la matrice valuée ci-dessus, nous construisons la matrice des relations entre régions et le graphe des collaborations intra et inter-régionales :

⁸Les indices a et b renvoient à des régions tandis que les indices k et k' renvoient à des agents (entreprises, centres de recherche,...).

	Région <i>a</i>	Région <i>b</i>	Région <i>c</i>	Région <i>d</i>	Région <i>e</i>
Région <i>a</i>	1	4	2	0	1
Région <i>b</i>	4	0	2	0	1
Région <i>c</i>	2	2	2	2	0
Région <i>d</i>	0	0	2	0	1
Région <i>e</i>	1	1	0	1	0

Matrice : Matrice évaluée des relations entre régions à partir de l'exemple 1.



Grphe 16 : Graphe évalué des relations entre régions.

Remarque : En général, lorsque l'on représente des réseaux de collaborations entre agents, on crée une matrice d'adjacence de ce réseau, X , définie de la manière suivante :

$$x_{ij} = \begin{cases} 1 & \text{si } i \text{ et } j \text{ ont au moins un projet en commun et si } i \neq j \\ 0 & \text{sinon} \end{cases}$$

avec x_{ij} qui rend compte de la relation entre l'agent j et l'agent i . Si nous avons retenu une matrice d'adjacence binaire, nous n'aurions pas pris en compte l'intensité des collaborations entre agents. Si les réseaux de NUTS étaient construits à partir de matrice d'adjacence binaire alors nous ne pourrions pas mesurer le nombre de collaborations totales entre deux régions mais seulement le nombre de collaborations non redondantes entre ces deux régions, c'est-à-dire le nombre de collaborations qui impliquent à chaque fois un binôme distinct. Dans l'exemple donné dans cette section, on s'aperçoit notamment que si l'on considérait uniquement une relation binaire, le graphe serait presque complet et il serait alors difficile de réaliser une analyse. En revanche, en prenant en compte (ce qui est le cas de l'étude) la valeur des liens on observe que la relation entre la région a et la région b est l'axe principal du réseau. Sur cet exemple, on observe clairement les informations supplémentaires que l'on peut obtenir grâce à l'intensité des liens par rapport à un cas binaire.

Après avoir présenté la méthode qui permet de représenter les réseaux de NUTS, nous appliquons cette méthode au domaine des Télécommunications et au domaine de l'Électronique-Microélectronique.

2.3.3 Des régions fortement connectées

Nous construisons les réseaux propres à chaque domaine technologique. Pour chacun des réseaux, nous étudions ses propriétés générales comme nous l'avons fait pour les réseaux de collaborations individuelles.

Afin de calculer les propriétés générales, nous nous ramenons à un réseau non valué. La valeur des liens influe peu sur les propriétés générales des graphes. Par conséquent, pour certains indicateurs comme la distance moyenne et le diamètre, il est presque équivalent de travailler avec un graphe binaire dont la valeur des liens est définie de la manière suivante : les liens entre NUTS prennent la valeur 1 si x_{ij}^R est supérieur ou égal à 1 et 0 sinon. Le tableau ci-dessous présente les propriétés générales pour les deux réseaux étudiés.

Indicateurs	<i>TEL</i> ¹	<i>EM</i> ²	<i>DDC</i> ³
Nombre de régions	189	179	211
Nombre de projets	187	167	354
Nombre de participations	2979	1969	4948
Nombre de liens	5916	5209	8198
Nombre potentiel de liens	17955	16110	22966
Densité (en %)	32,8%	32,3%	35,7%
Valeurs totales des liens	36074	17277	53351
Intensité moyenne des liens	6,1	3,3	6,5
Nombre de liens d'intensité 1 (en % du nombre de liens)	2134 36%	2196 42%	2847 35%
Nombre de liens réflexifs	89	80	113
Valeur des liens réflexifs	1196	550	1746
Nombre de composantes	1	1	1
Distance moyenne (<i>DM</i>)	1,68	1,69	1,64
Diamètre <i>D</i>	3	3	3
Degré moyen (Ecart-type)	375 (720)	190 (283)	497 (937)
Degré de centralité du graphe	8%	7,5%	7%

¹ *Tel* : Télécommunications ; ² *EM* : Electronique-Microélectronique ; ³ *DDC* : Les deux domaines confondus

Tableau 26 : Propriétés générales des réseaux de régions associés à chaque domaine technologique et au réseau des deux domaines confondus.

On observe que les deux réseaux sont de taille relativement similaire (si on se réfère au nombre de régions et au nombre de liens). De même, la répartition entre le nombre de liens de valeurs 1 et le nombre de liens supérieurs à 1 ainsi que le nombre de liens réflexifs sont relativement similaires. Toutefois, nous possédons une information supplémentaire qui est la valeur totale des liens. On s'aperçoit alors que le réseau des Télécommunications est construit à partir de deux fois plus de collaborations. Les relations entre régions sont plus denses dans le cas du réseau des Télécommunications puisque l'intensité moyenne des liens est de 6,1 contre 3,3 pour le réseau de l'Électronique Micro-électronique.

Lorsque l'on s'intéresse à la connexité du graphe ainsi qu'au diamètre et à la distance moyenne de celui-ci, on s'aperçoit que les deux réseaux possèdent des propriétés similaires. Les deux réseaux n'ont qu'une seule composante. De plus, la distance moyenne et le diamètre sont faibles (1,7 lien en moyenne

pour que deux régions soient liées). Le diamètre est quand à lui de 4. Cela signifie que toutes les régions présentes dans ces réseaux peuvent rapidement interagir avec d'autres régions présentes au sein des PCRD. Cette remarque est valable pour les régions des nouveaux États membres même si elles possèdent moins de liens. Toutes les régions font partie intégrante du réseau.

L'EER semble donc former un espace connexe et efficace dans la circulation des informations. Une étude complémentaire sur la localisation géographique ainsi que sur les distances moyennes entre agents permettrait de vérifier si la proximité sociale⁹ des régions coïncide avec une proximité géographique.

Nous avons vu que certaines régions occupent une place prépondérante dans les participations de leur pays. De même, il existe de fortes disparités entre les régions comme en témoignent les écarts-types élevés concernant le degré de ces régions. Dans le même temps, le degré de centralité du graphe est faible ce qui laisse supposer qu'il n'existe pas une seule région qui occupe une place centrale mais un sous-ensemble de régions. Dans la sous-section suivante, nous cherchons à déterminer le cœur des réseaux de régions ainsi que les axes de collaborations principaux.

2.3.4 Le cœur géographique du réseau européen : la prépondérance des régions des pays dominants

Il existe deux moyens d'identifier le cœur d'un réseau. Le premier consiste à représenter le graphe de chaque réseau et à ne conserver que les régions qui possèdent le plus de liens et les axes de collaborations entre les régions les plus intenses. Le second est celui que nous avons présenté à la section 1.5 qui est la méthode cœur/périphérie développée par Borgatti et Everett (1999).

Visualisation graphique des agents et des axes les plus importants

La difficulté réside ici dans la détermination des seuils pour juger qu'une région possède suffisamment de liens pour figurer dans le graphe et qu'un lien est suffisamment important pour être représenté. En couplant les axes les plus importants et les régions les plus connectées, il est ainsi possible de se faire une première idée du cœur du réseau.

Pour figurer dans le cœur du réseau, une région doit répondre à deux caractéristiques : posséder un nombre important de liens et être impliquée

⁹La proximité sociale correspond à la proximité des agents et/ou des régions au sein des réseaux sociaux, dans notre cas au sein des réseaux de collaborations en R&D.

dans un des axes de collaborations principaux du réseau. Dans la sous-section précédente, nous avons observé que le degré moyen de chaque réseau était très élevé avec un fort écart-type. Dans le même temps, nous avons remarqué que la centralité de degré du graphe était très faible. Ces deux mesures combinées signifient que chaque réseau compte de nombreuses régions avec beaucoup de liens. Nous retenons un second critère afin d'éviter d'obtenir un cœur trop élargi. De cette manière, nous mettons en avant à la fois les axes de collaboration principaux et les agents les plus centraux selon la mesure de degré. Étant donné que le degré est calculé dans le graphe valué, il est fortement probable que la plupart des axes de collaborations principaux relient deux régions qui possèdent de nombreux liens.

Il reste à déterminer le seuil à partir duquel nous considérons qu'un agent possède assez de liens pour figurer sur le réseau et le seuil à partir duquel, on considère que l'axe de collaboration fait partie des plus importants. Pour le premier, nous retenons le degré moyen de chaque réseau comme seuil, de même pour l'intensité des liens. Ainsi ne figurent dans le réseau Télécommunications que les régions qui ont un degré supérieur à 375 et qui possèdent au moins un lien d'intensité supérieure à 6. L'utilisation du degré moyen et de l'intensité moyenne permet ainsi de prendre en compte les différences de taille entre les deux graphes. Les graphes ainsi obtenus sont présentés en annexes A.10.1.

Sur ces deux graphes, on observe un grand nombre de régions encore présentes. Dans les deux cas, on dénombre 54 régions présentes dans les deux cœurs. Le réseau de chaque domaine technologique est difficilement lisible étant donné le nombre important de liens et d'agents. Afin d'identifier plus clairement quel est le cœur de chaque réseau, nous augmentons le seuil sur l'intensité des liens en conservant seulement les régions qui possèdent au moins un lien d'intensité supérieure à 50 pour le réseau Télécommunications et supérieure à 25 pour le réseau Electronique-Microélectronique. Ces deux seuils ont été obtenus par tâtonnement et nous conservons la proportionnalité entre les deux réseaux. Les graphes sont présentés dans l'annexe A.10.2.

Sur chacun de ces graphes, on observe que l'Ile de France est la région la plus centrale et possède un grand nombre de liens de forte intensité. Sur chaque graphe, on observe aussi clairement deux sous-ensembles de régions : le premier composé de régions qui sont reliées les unes aux autres et le second composé de régions qui possèdent uniquement un lien avec l'Ile de France. Sur ces deux graphes, l'Ile de France est le centre du réseau. Si l'on augmente de manière séquentielle le seuil concernant les liens, on observe qu'à un fort niveau d'intensité, il ne reste plus qu'une clique complète entre 5 régions.

Dans le cas du cœur du réseau Télécommunications, il existe quatre ré-

gions qui jouent un rôle de carrefour dans le réseau : Oberbayern, Angleterre Sud-Est, la Lombardie et la Communauté de Madrid. Mis à part pour Angleterre Sud-Est, toutes ces régions possèdent un lien réflexif qui est supérieur au seuil retenu ce qui signifie que les agents de ces régions collaborent de manière importante entre eux. Ceci peut apparaître en contradiction avec les conditions imposées par l'UE d'exiger au moins 3 agents de nationalités différentes au sein de chaque projet. Toutefois, les projets étant de taille largement supérieure à 3, il est tout à fait possible de collaborer en plus avec des agents du même pays voire de la même région. Dans ce graphe, on note la présence de la région Közép Magyaró (Hongrie) qui est la seule région d'un des nouveaux États membres. Dans le cœur étendu (annexes A.10.1), on en dénombrerait deux de plus : Mazowieckie et Malopolskie (Pologne). Ces trois régions correspondent aux principales villes de chaque pays à savoir Budapest pour la Hongrie et Varsovie et Cracovie pour la Pologne et correspondent surtout aux villes où sont localisés les agents les plus centraux qui appartenaient aux nouveaux États membres.

Dans le cœur du réseau Electronique-Microélectronique, on observe que le réseau est moins dense et compte moins de villes relais. Derrière l'Île de France, seule la région Oberbayern joue ce rôle. Ces deux régions et la région Rhône-Alpes sont les seules à conserver un lien réflexif. Dans ce graphe figure une seule région issue d'un des nouveaux États membres c'est la région polonaise de Mazowieckie (Varsovie). Dans le cœur étendu, on en comptait deux supplémentaires à savoir Prague (République Tchèque) et Közép Magyaró (Hongrie).

Si de nombreuses régions sont communes aux deux cœurs, certaines n'apparaissent que dans un seul. C'est le cas de la région de Prague qui ne figure que dans le cœur du réseau Electronique-Microélectronique. C'est le cas aussi de la région de Lisbonne qui n'apparaît que dans le cœur du réseau Télécommunications. Ces régions cohabitent avec des régions d'envergure supérieure qui sont des régions centrales dans les deux domaines.

Dans le paragraphe suivant, nous vérifions les résultats obtenus à l'aide des graphes grâce à la méthode cœur/périphérie.

La méthode cœur/périphérie

Puisque le réseau est valué, il apparaît pertinent d'utiliser comme mesure de fitness la densité plutôt que le coefficient de corrélation. Pour chaque domaine technologique, nous déterminons donc le cœur du réseau de régions. Afin d'être certain que le résultat obtenu est bien le meilleur possible et non pas un maximum local, nous utilisons l'algorithme proposé par Borgatti et

Everett à partir de plusieurs points de départ différents obtenus en permutant certaines lignes de la matrice. Il est important de souligner que les liens réflexifs ne rentrent pas en compte dans le calcul de la densité.

Pour le réseau Télécommunication, l'ensemble des régions qui composent le cœur est donné dans l'annexe A.11.1. Le cœur est constitué de 72 régions (sur 189). L'ensemble des régions que nous avons identifié comme appartenant au cœur au paragraphe précédent se retrouvent dans le cœur identifié par la méthode cœur/périphérie. La matrice de densité est la suivante :

Télécommunications	Cœur	Périphérie
Cœur	5,228	0,323
Périphérie	0,323	0,055

La densité entre régions du cœur est largement supérieure à celle des relations dans les trois autres blocs. La partition fournit par l'algorithme semble cohérente. Dans le cœur apparaissent deux nouvelles régions des nouveaux Etats membres : la région de Bucarest (Roumanie) et la Slovénie.

Pour le réseau Electronique-Microélectronique, nous avons rencontré une difficulté. Quelle que soit la matrice donnée en entrée, l'algorithme utilisant comme mesure de fitness la densité renvoie toujours la même partition à savoir un seul élément dans le cœur : la région Ile de France. Ce phénomène s'explique en grande partie par la prépondérance de cette région dans le réseau comme nous l'avons vu au paragraphe précédent. Afin de contourner cette difficulté, nous retenons à la place une mesure de fitness différente à savoir le coefficient de corrélation. Les régions présentes dans le cœur sont données dans l'annexe A.11.2. La mesure de fitness et la matrice de densité sont données ci-dessous :

Mesure de fitness : 0,641

Electronique-Microélectronique	Cœur	Périphérie
Cœur	14,337	2,146
Périphérie	2,146	0,352

Le cœur du réseau Electronique-Microélectronique est de taille beaucoup plus réduit que celui du réseau Télécommunications. Comme pour le réseau Télécommunications, la densité du cœur est bien plus élevée que celle des relations entre régions de la périphérie. On observe que la densité du cœur du réseau Electronique-Microélectronique est bien supérieure

à celle du réseau Télécommunications. Parallèlement, la densité des relations entre régions du cœur et de la périphérie est aussi plus élevée dans le réseau Electronique-Microélectronique. Ceci signifie donc que le réseau Electronique-Microélectronique possède certes un cœur plus restreint mais qu'il existe des relations de collaborations importantes entre régions du cœur et de la périphérie alors que dans le réseau Télécommunications, les relations les plus importantes font partie du cœur.

Remarque : Si pour le réseau Télécommunications, nous utilisons comme mesure de fitness le coefficient de corrélation et non la densité, il faut noter que la partition que nous renvoie l'algorithme est strictement identique.

Lorsque l'on observe les deux cœurs, la première remarque que l'on peut faire est que les régions du cœur du réseau Electronique-Microélectronique font toutes parties du cœur du réseau des Télécommunications à l'exception de la région Rhône-Alpes. Ces régions semblent donc constituer le socle de la recherche en Europe du moins dans les deux domaines technologiques étudiés. La seconde remarque concerne les régions des nouveaux Etats membres. Elles sont peu nombreuses dans les deux cœurs (une dans le réseau Electronique-Microélectronique et trois dans le réseau Télécommunications) et correspondent aux régions où sont situés les agents les plus centraux.

Les deux sous-sections suivantes vont nous permettre d'affiner à la fois la définition des régions motrices de l'Europe et la place qu'occupent réellement les régions issues des nouveaux États membres par l'intermédiaire des indicateurs de centralité utilisés jusqu'à présent.

2.3.5 Conformité des PCRDT à la géographie de l'innovation européenne

Il est intéressant, à partir des indicateurs réseaux que l'on a présentés, de chercher à établir quelles sont les régions les plus performantes en matière de R&D. A partir des données dont nous disposons, le critère de performance est déterminé à partir de variables liées à la participation des agents des différentes régions. Il serait intéressant de disposer de données concernant les débouchés des collaborations dans le cas des PCRDT pour compléter l'analyse.

Cette comparaison peut être rendue plus intéressante encore si nous prenons comme point de repère les statistiques établies par Eurostat et Inno Metrics (deux organismes liés et dépendants de la Commission Européenne).

En effet, il nous sera possible alors de déterminer si au sein des PCRDT, du moins au sein des deux domaines technologiques étudiés, les régions les plus performantes sont les mêmes que celles mises en avant par Eurostat et Innometrics. De plus, cette comparaison nous permettra d'étudier la position occupée par les nouveaux États membres. Il serait logique que les régions des nouveaux États membres soient mieux placées dans les PCRDT que dans les statistiques d'Eurostat étant donné l'existence de dispositifs incitatifs pour l'intégration des nouveaux États membres dans les PCRDT. La comparaison est toutefois rendue délicate par le fait que le panel d'Eurostat est composé uniquement de régions européennes alors que les PCRDT s'adressent aussi à des pays hors Europe et ne contient pas toutes les régions européennes mais seulement celles qui compte au moins une participation.

Des indicateurs de participation absolus

Pour déterminer quelles sont les régions leaders dans les PCRDT, nous utilisons trois critères :

- Le nombre d'agents de chaque région qui participent aux PCRDT. Cet indicateur renseigne sur la capacité d'une région à intégrer plusieurs de ses agents dans les programmes cadres et donc sur son potentiel de recherche ;
- Le nombre de participations de ces agents. S'il est important qu'une région possède de nombreux agents dans les PCRDT, il faut aussi que ceux-ci y jouent un rôle actif et qu'ils intègrent plusieurs projets ;
- Le nombre de liens des agents ; Cet indicateur permet de prendre en compte indirectement l'envergure des projets auxquels les agents de la région participent et aussi l'intensité de leur collaboration.

Ces trois indicateurs sont liés. Toutefois, établir un classement ou déterminer la région leader à partir d'un seul de ces critères pourrait nous exposer à des biais. Par exemple, une région peut compter un nombre très important d'agents qui participent très peu et nouent peu de liens au sein des réseaux. Utiliser le nombre d'agents comme unique critère reviendrait alors à surévaluer la place de cette région. Aucun de ces trois critères n'a une importance supérieure à un autre c'est pourquoi nous n'appliquons aucune pondération. Nous retenons, comme pour les agents, une moyenne harmonique pour les mêmes raisons qu'évoquées précédemment.

Nous calculons ces trois critères pour chaque région et nous calculons par la suite le rang moyen de chaque région pour chaque domaine technologique ainsi que pour le réseau agrégé (les tableaux 1, 2 et 3 de l'annexe A.12 donnent le classement des 20 régions leaders de chaque réseau). Les régions leaders de

chaque domaine technologique font toutes partie du cœur du réseau identifié à la sous-section précédente. Les régions dont les agents sont les plus impliqués sont celles des grandes capitales économiques européennes comme l'Ile de France, Oberbayern, la Communauté de Madrid, la Lombardie, . . . Parmi les régions les plus participatives, on retrouve un grand nombre de régions issues des cinq pays majeurs des PCRDT que sont l'Allemagne, la France, l'Italie, le Royaume-Uni et l'Espagne ainsi que des régions des pays nordiques et des pays du Benelux qui font partie des pays les plus innovants selon Eurostat (cf. tableau 5 en annexes A.12). Si l'on compare les régions les plus participatives que nous avons identifiées dans les deux domaines technologiques avec les régions qui dépensent le plus en R&D et qui possèdent le plus de personnel travaillant en R&D tous secteurs confondus, nous nous apercevons que de nombreuses régions sont communes. Ceci signifie que de manière générale, les PCRDT reflètent la géographie de l'innovation en Europe avec les mêmes régions fortes.

Il faut noter qu'il existe des différences entre les deux domaines technologiques même si ces deux domaines ont 13 régions en commun parmi les 20 régions les plus participatives. Certaines régions possèdent un rang bien meilleur dans un domaine plutôt que dans un autre. Ainsi, la région Rhône-Alpes est classée 3^{ème} pour le réseau Electronique-Microélectronique alors qu'elle ne figure pas parmi les 20 premières du réseau Télécommunications (58^{ème}). De la même manière, la région de Lisbonne est mieux classée dans le réseau Télécommunications (15^{ème}) que dans le réseau Electronique-Microélectronique (70^{ème}). Il existe donc des effets de spécialisation d'un domaine à l'autre au sein des PCRDT même si ces effets demanderaient à être confirmés dans le temps. De plus, il est difficile de savoir si ces régions utilisent les PCRDT pour rattraper leur retard dans des secteurs où elles sont peu avancées ou s'il s'agit d'un vrai effet de spécialisation déjà présent. Pour pouvoir déterminer la raison de cette spécialisation, il faudrait croiser nos bases de données avec des bases de données autres que celles des PCRDT afin de connaître les domaines de référence de chaque région.

La dernière remarque qui provient de ces tableaux c'est l'absence quasi totale des régions des nouveaux Etats membres. Seule la région de Közép Magyarorszag est présente dans les régions les plus participatives du domaine Télécommunications. Les tableaux ci-dessous donnent le rang des 5 régions des nouveaux États membres les mieux classées pour chaque réseau :

Rang	Télécommunications
19	hu10 Közép-Magyarország
43	pl12 Mazowieckie
52	si00 Slovénie
61	cz01 Praha
63	pl21 Malopolskie

Rang	Electronique-Microélectronique
24	pl12 Mazowieckie
34	hu10 Közép-Magyarország
56	cz01 Praha
57	ro Bucuresti-Ilfov
59	lt00 Lituanie

Rang	Les deux domaines confondus
22	hu10 Közép-Magyarország
33	pl12 Mazowieckie
62	cz01 Praha
63	lt00 Lituanie
68	ro Bucuresti-Ilfov

Tableaux 27, 28 et 29 : Les régions des nouveaux États membres les mieux classées dans les différents domaines technologiques.

Les régions les mieux classées parmi les nouveaux États membres sont celles qui ont été identifiées à la section précédente comme possédant un agent particulièrement actif dans un domaine technologique, agent qui figure notamment parmi les agents les plus centraux. Il peut apparaître normal que les régions des nouveaux États membres soient assez mal classées lorsque l'on utilise des indicateurs de participation absolue qui ne prennent pas en compte la taille des régions (en matière de R&D) notamment. Dans le paragraphe suivant, nous tentons de prendre en compte cette dimension.

Des indicateurs de participation relatifs

Afin d'avoir une idée relative de la performance des régions au sein des PCRDT, il est donc important de prendre en compte les moyens dont elles disposent.

Pour évaluer la place des régions au sein des PCRDT, nous retenons deux critères.

Premier critère, la participation moyenne d'une région PM_R qui est égale à :

$$PM_R = \frac{\text{Nombre de participations des agents de la région}}{\text{Nombre d'agents de la région}}$$

Cet indicateur permet ainsi d'évaluer au sein des PCRDT l'implication des agents de chaque région. Nous calculons cet indicateur pour chaque région et pour chaque domaine technologique ainsi que pour le réseau agrégé. Pour éviter que des régions possédant un faible nombre d'agents possèdent un indicateur de participation moyen élevé, nous fixons un seuil minimum de trois agents pour une région dans un domaine technologique pour calculer cet indicateur. De fait, de nombreuses régions localisées dans les nouveaux États membres ne peuvent être classées faute d'un nombre d'agents suffisant.

Second critère, un indicateur de performance relatif (IP_R) qui prend en compte la taille des régions à l'aide d'une variable extérieure à notre base de données. Cet indicateur est égal au rapport entre le nombre de liens cumulés des agents de chaque région et les effectifs de R&D de la région en effectif équivalent temps plein (ETP) (Source : Eurostat).

$$IP_R = \frac{\text{Nombre de liens cumulés des agents de la région}}{\text{Nombre de chercheurs en ETP}}$$

Cet indicateur repose sur l'hypothèse que chaque lien demande un investissement en matière de collaboration et qu'il faut donc disposer d'une main d'œuvre suffisante et qualifiée pour pouvoir s'impliquer de manière efficace dans toutes les collaborations. Les effectifs correspondent à un effectif moyen sur la période 2002-2006. Nous ne disposons ni de toutes les années, ni d'une année commune à chaque région. Nous nous sommes assurés que nous disposions des données pour les principales régions identifiées jusqu'alors ainsi que pour une grande partie des nouveaux États membres dont les plus participatifs (sauf PL21 Molopolski). Certains pays ne sont pas renseignés ou peu comme la Suisse et la Belgique ainsi que l'ensemble des pays non européens. Le nombre de régions renseignées est le suivant :

Domaine technologique	Nombre de régions	Régions renseignées	Régions comptant plus de trois agents
Télécommunications	189	149	97
Electronique-Micro.	179	144	95
Réseau agrégé	211	165	124

Tableau 30 : Les régions dans les différents domaines technologiques.

De la même manière qu'au paragraphe précédent, nous calculons un rang moyen (moyenne harmonique) à partir de ces deux critères.

Lorsque l'on observe quelles sont les régions les plus participatives à l'aide des indicateurs relatifs (tableaux 1 bis, 2 bis et 3 bis en annexes A.12), un certain nombre de nouvelles régions apparaissent pour chaque domaine technologique. Ces régions sont des régions de taille plus réduite en termes de moyens en R&D. Par exemple, la région de Nord-Brabant (Belgique) dans le domaine Electronique-Microélectronique se caractérise par un taux de participation moyen très élevé de 5,57 (39 participations pour 7 agents différents) et par un indicateur de performance assez élevé de 0,03. Ces classements montrent que les régions les plus participatives dans l'absolu ne le sont pas forcément de manière relative. C'est le cas notamment de l'Ile de France qui, si elle figure dans le classement du réseau Télécommunications, ne figure pas dans celui du réseau Electronique-Microélectronique ni dans celui du réseau agrégé, au contraire de la région Oberbayern qui est bien classée au regard des indicateurs absolus comme des indicateurs relatifs.

Malgré le fait que l'on ait pris en compte un effet taille, on observe peu de régions issues des nouveaux États membres qui émergent. La région de Prague apparaît dans le classement du réseau Electronique-Microélectronique. La région de Chypre est présente dans le classement du réseau agrégé. Si de nombreuses régions, qui n'apparaissent pas à partir des indicateurs absolus, se signalent maintenant à partir d'indicateurs relatifs, ce n'est pas le cas des régions des nouveaux États membres. Les tableaux ci-dessous donnent le rang des 5 régions localisées dans les nouveaux États membres les mieux classées pour chaque secteur :

Rang	Télécommunications
40	ro Bucuresti-Ilfov
41	pl12ãMazowieckie
42	hu10ãKözép-Magyarország
52	pl41ãWielkopolskie
59	cz01ãPraha

Rang	Electronique-Microélectronique
13	cz01Prah
42	hu10Közép-Magyarország
54	ro Bucuresti-Ilfov
86	sk01Bratislavský kraj
88	pl21Malopolskie

Rang	Les deux domaines confondus
18	Cypre
21	pl21Malopolskie
37	hu10Közép-Magyarország
44	pl12Mazowieckie
49	pl41Wielkopolskie

Tableaux 31, 32 et 33 : Les régions des nouveaux États membres les mieux classées dans les différents domaines technologiques.

Malgré le faible nombre de régions classées, ces régions occupent un rang assez faible. Elles se comportent mieux lorsque le réseau est agrégé. Ces régions correspondent pour la plupart à des régions où nous avons identifié un agent central comme la région Közép-Magyarország (Hongrie) ou encore la région de Mazowieckie (Pologne). Si l'on compare ces résultats avec ceux d'Eurostat, on observe que des régions localisées dans les nouveaux États membres étaient très bien classées selon l'indicateur synthétique d'Eurostat pour 2006. Ainsi, la région de Prague se classe 15^{ème}, la région de Bratislava 27^{ème} et la région de Közép Magyarország 34^{ème}. Si l'on retrouve bien ces régions parmi les régions des nouveaux États membres les plus participatives, il faut constater qu'elles sont moins bien classées en termes de participation qu'elles ne le sont par Eurostat de manière globale à partir d'un nombre de critères plus hétérogène et plus global. Cette différence peut s'expliquer de deux manières. La première concerne les thématiques financées par l'UE qui peuvent ne pas correspondre à la spécialité de ces pays notamment la thématique IST étudiée dans la thèse. De fait, les agents de ces pays ont un intérêt moindre à participer. La deuxième explication peut venir d'une difficulté à s'intégrer aux réseaux des PCRDT, au delà d'une simple participation, et à intégrer notamment les projets d'envergure. Malgré le dispositif incitatif et la volonté de l'UE d'intégrer à l'EER les agents des nouveaux États membres, ces agents peuvent ne pas occuper des places significatives au sein du réseau et intégrer un nombre de projets plus faibles que les régions de l'UE15.

Nous aurions pu retenir comme critère de taille les dépenses en R&D mais le nombre de données manquantes était supérieur à celui obtenu pour

les effectifs en ETP. Dans notre base de données, nous ne possédions aucune donnée précise sur les effectifs engagés et les données sur les coûts supportés par les agents et les financements reçus ne sont pas complètes. La principale limite de notre indicateur c'est que le dénominateur est identique quel que soit le domaine technologique alors qu'il aurait été préférable d'avoir une décomposition par secteur pour prendre en compte des effets de spécialisation au sein des régions.

Les indicateurs réseaux ont permis de mettre en avant les régions motrices au sein des PCRDT. Ces régions sont les mêmes régions que l'on peut retrouver dans les études d'Eurostat sur l'innovation de manière globale. Les régions localisés dans les nouveaux États membres ne se distinguent ni à partir des indicateurs de participation absolus ni à partir des indicateurs de participation relatifs. Malgré les conditions imposées par l'UE et les incitations pour favoriser la participation des agents localisés dans les nouveaux États membres, nous n'observons pas un phénomène de rattrapage entre ces États et ceux de l'UE à 15 membres. On observe certes une hausse de la participation, certainement due aux conditions imposées, mais pas de réel rattrapage en matière de place occupée au sein du réseau et de prépondérance dans les réseaux de collaborations. Le 6^{ème} PCRDT est le premier à se dérouler dans le cadre de l'intégration des nouveaux États membres. De ce point de vue, l'augmentation du nombre de participations des agents localisés dans les nouveaux États membres, même si elle ne s'accompagne pas d'agents centraux, peut donc être vue comme un premier pas dans une volonté de rattrapage. Les collaborations au sein du 7^{ème} PCRDT permettront de vérifier si ce phénomène se poursuit.

Conclusion

L'objectif de cette partie était de montrer dans quelle mesure l'analyse réseau permet de mieux appréhender la structure des collaborations au sein des PCRDT. Tout au long de cette partie, nous avons pu remarquer que l'analyse réseau permettait de vérifier partiellement si les objectifs assignés au PCRDT ont été remplis. L'analyse réseau peut être utilisée comme un outil contribuant à l'évaluation des politiques économiques du type de celles des PCRDT qui visent à intensifier les collaborations entre agents. L'analyse réseau avait d'ailleurs été utilisée par A. Tholoniât (2008) comme un outil complémentaire dans la mise en place d'une démarche d'évaluation des pôles de compétitivité.

Pour rappel, les PCRDT ont pour objectif la création d'un EER qui se veut un espace de collaboration entre agents de nature différente. Au travers le financement de projet en R&D, l'UE cherche à inciter les agents à collaborer afin d'atteindre une masse critique et de partager leurs connaissances et compétences. L'ensemble de ces collaborations doit former un réseau dense où les agents interagissent entre eux pour générer de nouvelles connaissances et des externalités favorables à l'innovation. De cette manière, l'UE espère ainsi améliorer la compétitivité de son secteur recherche, de ses entreprises et, de manière, plus globale rattraper son retard vis-à-vis du Japon et des Etats-Unis. Les PCRDT sont le principal moyen mis en œuvre par l'UE pour favoriser l'innovation en Europe.

L'analyse réseau nous a permis d'étudier les propriétés générales des graphes formés par les secteurs Electronique-Microélectronique et Télécommunications. Cette étude a notamment mis en avant que le réseau de collaborations des PCRDT possédait des propriétés favorables à l'innovation. Le réseau de chaque domaine technologique est composé à chaque fois d'une grande composante qui comprend plus de 95% des agents du réseau (100% même dans le réseau Electronique-Microélectronique). De plus, ces compo-

santes se caractérisent par des distances moyennes relativement faibles et un diamètre peu élevé. Les agents peuvent donc communiquer et échanger entre eux par un faible nombre d'intermédiaires. Ces propriétés sont d'autant plus importantes que des phénomènes de "decay" ou de réseau borné existent au sein du réseau (si ce n'était pas le cas, on pourrait alors supposer qu'une seule participation serait suffisante pour accéder à l'ensemble des ressources du réseau et donc, de manière rationnelle, l'agent s'en contenterait ce qui n'est pas le cas dans la réalité). Ainsi, il est important pour les agents d'être situés à proximité afin de pouvoir profiter des connaissances et des externalités générés par les autres projets. Nous avons cherché à comparer les réseaux avec des "small world networks" qui sont des réseaux qui possèdent des propriétés susceptibles d'être propices à l'innovation. Si les réseaux des PCRDT ne possèdent pas encore toutes les propriétés d'un "small world networks", ils possèdent toutefois des propriétés proches. De plus, ils présentent l'avantage d'être plus robustes que ces derniers à la suppression des agents centraux.

Nous avons montré que l'analyse réseau permettait d'étudier l'évolution des réseaux et donc de comparer deux programmes cadres. Les problèmes de correspondance entre les PCRDT rendent ardues les comparaisons temporelles. Toutefois, nous avons utilisé le secteur des nanotechnologies afin d'avoir au moins un aperçu de l'évolution entre les programmes cadres. Lorsque l'on étudie l'évolution des propriétés du réseau, on remarque que le réseau tend de plus en plus vers un espace intégré avec une seule composante caractérisée par une faible distance entre les agents. De plus, celui-ci devient plus robuste à la suppression des agents centraux signe d'indépendance par rapport aux hubs des réseaux. Bien que le secteur choisi soit favorable à une telle évolution puisqu'il s'agit d'un secteur en développement actuellement, on remarque que le nombre d'agents, de liens, . . . a évolué plus que proportionnellement par rapport au budget proposé, signe d'une véritable intensification des échanges.

L'UE a réussi à attirer au sein des PCRDT des agents leaders dans leur domaine notamment des grands groupes industriels et d'importants centres de recherche. Ces agents constituent même des agents centraux des différents réseaux signe d'une véritable implication et donc d'une réelle attractivité des PCRDT en matière d'innovation. De plus, ces agents sont présents d'un programme à l'autre. Cela permet d'assurer une continuité entre les deux programmes cadres et ainsi d'attirer de nouveaux agents pour le programme cadre suivant grâce à l'image de prestige ainsi renvoyée.

Au sein des réseaux, nous retrouvons une diversité d'agents aussi bien en termes de type que de nationalité, diversité souhaitée par l'UE afin de favoriser les collaborations entre agents différents ainsi que l'intégration des nouveaux États membres. Toutefois, si ces agents appartiennent au réseau,

ils n'y jouent pas un rôle majeur. Ainsi s'agissant des PME, leur nombre a fortement augmenté d'un programme à l'autre mais sans se traduire par un meilleur positionnement au sein du réseau comme en témoignent les indicateurs de centralité moyens. Concernant les agents des nouveaux États membres, le constat est équivalent. Ils appartiennent au réseau mais peu d'entre eux sont des agents centraux.

Au niveau des régions, il s'avère que les régions localisées dans les nouveaux États membres comptabilisent au final un faible nombre de participations et que, même de manière relative, leur participation reste en retrait. On retrouve les principales régions leaders dans le domaine de la recherche en Europe comme le montre les comparaisons avec les indicateurs d'Eurostat. A défaut d'un développement harmonieux des activités de R&D, on est plutôt dans un schéma de reproduction de l'existant.

Ces conclusions restent valables seulement pour les domaines étudiés. Il faudrait pouvoir vérifier cela pour l'ensemble des programmes cadres. Une comparaison temporelle entre le 5^{ème} PCRDT (avant le début du processus d'intégration), le 6^{ème} (pendant) et le 7^{ème} (après) apporterait des informations supplémentaires. L'analyse réseau serait alors un soutien pour évaluer la manière dont les agents des nouveaux États membres s'intègrent au sein des PCRDT.

L'analyse réseau a permis de souligner le faible rôle joué par les agents que l'UE souhaite voir participer aux PCRDT. Il est difficile de contrôler si ces agents, en plus de participer faiblement aux PCRDT, ne servent pas d'agents alibis, c'est-à-dire que la participation de ces agents seraient une condition nécessaire et favorable pour voir un projet financé. La vérification d'une telle hypothèse est ardue. Deux moyens peuvent être envisagés pour cela. Le premier consisterait à construire les réseaux de collaborations en tenant compte des financements obtenus et/ou des coûts supportés au sein de chaque projet. De cette manière, ne figurent sur le réseau que les agents dont l'implication est jugée suffisante. La principale difficulté de cette approche est de déterminer un seuil pour les participations suffisantes étant donnée la diversité des agents et des instruments. De plus, nous nous heurtons aussi à des problèmes liés au remplissage de la base de données. Les champs de financement et de coût individuel étant deux des champs les plus mal renseignés. Le second moyen consiste à utiliser l'économétrie afin de déterminer les facteurs favorables au financement des projets, le nombre de PME et/ou d'agents localisés dans les nouveaux États membres étant alors utilisé comme variables explicatives. Il faudrait pour cela connaître l'ensemble des projets déposés auprès de la Commission Européenne, l'ensemble des critères retenus pour la sélection ainsi que les notes attribuées sur les critères scientifiques.

L'analyse des réseaux sociaux permet de déterminer, en partie, dans quelle mesure les objectifs assignés aux PCRDT en matière de collaboration en R&D ont été atteints. De plus, l'analyse des réseaux apporte des éléments supplémentaires sur la caractérisation des réseaux et des agents. Cependant, cette approche ne permet pas de vérifier l'ensemble des objectifs assignés aux PCRDT. En effet, l'analyse des réseaux ne fournit aucune information sur la performance réalisée en termes d'innovation. Pour cela, une possibilité serait de croiser nos données avec des bases relatives aux dépôts de brevets par exemple afin d'observer si les agents qui participent aux PCRDT voient s'accroître le nombre de brevets déposés et si les régions les plus participatives sont aussi celles d'où émanent le plus de brevets. Une des difficultés serait alors de distinguer la performance liée aux PCRDT et celle liée à d'autres programmes de recherche ou à l'activité normale de chaque agent et de chaque région. Il pourrait être intéressant de voir si les régions les plus collaboratives des PCRDT sont aussi celles dans lesquelles on retrouve le plus grand nombre de brevets même s'il sera alors difficile d'identifier dans quel sens joue la relation. Cette approche est rendue difficile pour deux raisons. La première c'est le décalage temporel entre les collaborations et le dépôt de brevet. La seconde c'est le fait que les brevets ne constituent pas forcément l'unique moyen de mesurer la performance d'une collaboration ou de manière plus générale la performance en R&D.

Les indicateurs réseaux ne permettent pas d'étudier la réussite ou non des collaborations en matière d'innovation. Certains articles cherchent à déterminer les facteurs clés du succès ou de l'échec des collaborations (Powell, Koput, Smith-Doer et Owen-Smith, 1999 ; Goerzen, 2007). Dans cette approche, on cherche aussi à montrer que la place des agents dans les réseaux de collaborations peut être favorable à l'innovation au niveau individuel. Cette approche est la moins développée car il existe de réelles difficultés à mesurer la réussite ou non d'une alliance. Deux problèmes majeurs apparaissent. Premièrement, comment mesurer la performance des agents dans une alliance en termes d'innovation et, deuxièmement, comment distinguer dans cette performance la partie exclusivement due au fait d'avoir collaboré ? Le plus souvent, les études portent sur les brevets. On vérifie notamment si des collaborations à une date t ont donné lieu à des dépôts de brevets dans le futur. Dans cette approche, le réseau n'est plus le phénomène que l'on cherche à étudier, il devient un des facteurs susceptibles d'expliquer le dépôt de brevets par plusieurs agents. De manière complémentaire, on peut s'intéresser à l'impact des réseaux de collaborations non pas sur un agent en particulier mais sur la performance d'un pays ou d'un ensemble de pays en matière d'innovation. Il s'agit de déterminer quel réseau de collaboration est nécessaire afin

d'innover de la manière la plus efficace possible. Dans le cas des PCRDT, une telle approche pourrait être utilisée afin d'observer si les collaborations à une date t dans les PCRDT peuvent influencer le dépôts de brevets ou les co-publications en $t + 1$ par exemple.

Les réseaux et les indicateurs mis en place pour les étudier peuvent servir dans une seconde étape à observer si les agents ou les régions mis en évidence grâce à l'analyse réseau sont par la suite des agents ou des régions dont l'activité d'innovation est favorisée. En elle-même l'analyse réseau ne permet pas de mesurer l'innovation, elle fournit des indicateurs complémentaires pour une évaluation plus précise des PCRDT.

Dans cette partie, nous avons notamment montré qu'une partie des agents présents dans le 5^{ème} PCRDT dans le sous-domaine des Nanotechnologies était à nouveau présent dans le 6^{ème} PCRDT. Il existe donc des incitations pour un agent à continuer de participer dans le programme cadre suivant. Dans la partie suivante, nous nous intéressons aux déterminants qui incitent les agents, qui ont déjà participé au 5^{ème} PCRDT, à collaborer les uns avec les autres dans le 6^{ème} PCRDT afin d'expliquer en partie la formation des réseaux. Nous montrerons que l'analyse réseau est utile pour montrer que les collaborations dans le 5^{ème} PCRDT influent sur les collaborations formées dans le 6^{ème} PCRDT. Les indicateurs réseau ne sont plus au centre de l'analyse mais ils sont utilisés comme variables explicatives supplémentaires dans les modèles économétriques examinant les déterminants des collaborations en R&D dans le cadre des PCRDT. Nous vérifions si la place occupée par les agents dans le 5^{ème} PCRDT influe sur les choix de collaborations. Ainsi, dans la partie suivante, nous expliquons la formation des collaborations entre agents. Parmi les déterminants que nous testons figurent des déterminants traditionnels comme les moyens en R&D ou encore la proximité géographique auxquels nous ajouterons des déterminants liés aux réseaux, plus précisément des déterminants liés à la proximité sociale entre agents et au prestige acquis par les agents participant au 5^{ème} PCRDT.

Deuxième partie

Déterminants des choix de collaborations au sein des PCRDT

Introduction

Dans la première partie de la thèse, nous avons montré de quelle manière l'analyse réseau pouvait permettre de mieux appréhender les réseaux de collaborations formés au sein des PCRDT en fournissant un certain nombre d'indicateurs susceptibles d'améliorer la connaissance de la structure du réseau et des principaux agents qui y participent. Il s'agit de l'approche la plus courante pour aborder les réseaux et plus particulièrement ceux qui découlent des collaborations au sein des programmes cadres. Nous avons montré que l'analyse réseau pouvait permettre d'évaluer, partiellement, la politique des PCRDT menée par l'Union Européenne.

En revanche, nous n'avons pas étudié la source même de ces réseaux, c'est-à-dire les choix collaboratifs des agents. Plus précisément, jusqu'à présent nous nous sommes intéressés essentiellement à la structure du réseau une fois celui-ci formé, mais nous ne nous sommes pas intéressés aux facteurs qui affectent le choix de partenaires et qui ont contribué à former les réseaux étudiés dans la partie I. Ces choix constituent pourtant une étape décisive puisqu'ils conditionnent au niveau individuel la réussite du projet collaboratif de l'agent et, surtout, ce sont ces choix mis bout à bout qui conduisent à la formation des véritables réseaux de collaborations. L'économétrie de réseau peut permettre d'estimer quels sont les déterminants pertinents pour expliquer les choix collaboratifs. Cet outil fait partie de l'analyse réseau. En effet, afin de montrer l'influence des réseaux établis sur les choix de collaborations futurs, nous mobiliserons une partie des indicateurs présentés à la partie I.

Plusieurs études ont cherché à identifier les facteurs susceptibles d'expliquer que les agents collaborent les uns avec les autres en matière de R&D, que ce soit au niveau théorique (D'Aspremont and Jacquemin, 1988 ; Cohen and Levinthal, 1990 ; Cassiman and Veugelers, 2002) ou au niveau empirique (Autant-Bernard et al., 2007 ; Arranz et Fernandez de Arroyabe, 2008 ; Hanaki, Nakajima et Ogura, 2007 ; Paier et Scherngell, 2008). D'un point de vue théorique, D'Aspremont et Jacquemin (1988) ont ainsi montré dans le cas particulier du duopole que deux agents pouvaient avoir intérêt à collabo-

rer afin de partager les coûts liés à l'investissement en R&D avant de se faire concurrence sur le marché. Le partage des coûts liés à la recherche mais aussi le partage des risques, la capacité des agents à s'approprier les connaissances, la pluri-disciplinarité de la recherche, la complémentarité entre les agents et les externalités de connaissance sont autant de facteurs incitatifs qui sont traditionnellement mis en avant pour expliquer les choix de collaboration.

Deux autres déterminants sont venus s'ajouter à cette liste au fil des différents travaux menés par les économistes. Il s'agit des facteurs liés aux réseaux préexistants et de la proximité géographique. Un grand nombre d'études a mis en avant le rôle joué par les réseaux antérieurs pour expliquer les choix de collaborations que ce soit au travers de la position occupée par l'agent au sein de ce réseau, par des collaborations passées ou encore par l'existence de partenaires communs (Gulati, 1995, 1998 ; Bala et Goyal, 2004 ; Fafchamps, Goyal et van der Leij, 2010). On parle alors d'effets réseaux pour expliquer les choix de collaborations. Le second déterminant susceptible d'affecter les choix de collaboration et qui connaît un intérêt croissant est la proximité géographique. De récentes études ont montré que la proximité géographique pouvait avoir un impact sur l'innovation notamment en favorisant les contacts et les relations entre agents (Zucker and al., 1994 ; Almeida and Kogut, 1999 ; Singh, 2005 ; Breschi and Lissoni, 2009). De manière plus nuancée, Ponds, van Oort et Frenken (2007) montrent que la proximité géographique peut jouer différemment selon le type des agents. Ainsi, les auteurs mettent en avant le fait que la proximité géographique semble jouer un rôle plus important pour les collaborations entre agents académiques et agents non académiques que pour les collaborations entre agents académiques seulement.

La principale difficulté et limite relative à l'étude de ces choix au niveau des PCRDT réside dans l'hypothèse initiale que nous avons présenté dans la première partie de la thèse concernant la construction du réseau. En effet, nous avons supposé que les agents qui participaient à un même projet possédaient tous un lien bilatéral entre eux. Cette hypothèse est discutable car, lorsque les agents participent aux PCRDT, ils soumettent un projet à la Commission européenne qui est porté par un consortium d'agents. C'est pourquoi de nombreux auteurs s'intéressent à des collaborations qui concernent relativement peu d'agents comme les contrats qui relient les agents deux à deux ou les réseaux de co-auteurs.

Fafchamps, Goyal et van der Leij (2010) s'intéressent ainsi aux réseaux de co-auteurs en économie. Cette discipline présente l'avantage que le nombre de co-auteurs par papier est réduit (le plus souvent deux à quatre co-auteurs) contrairement à des disciplines comme la médecine ou la physique où peut

figurer sur les papiers l'ensemble des chercheurs d'un laboratoire. Ces auteurs mettent notamment en avant l'importance des collaborations antérieures à travers le nombre des publications passées (comme signe de prestige et de notoriété) pour expliquer les collaborations à une période donnée.

Il peut paraître étrange d'étudier la formation de liens bilatéraux alors que les agents s'engagent à collaborer non pas avec un agent en particulier mais avec un sous-ensemble défini d'agents. Il serait donc peut être plus opportun de chercher à déterminer les facteurs qui incitent un agent à faire partie d'un projet. Pour autant, on peut aussi supposer que si un agent participe à un projet c'est qu'il compte tirer profit de celui-ci mais aussi de la collaboration qu'il mettra en œuvre avec chaque membre du projet. En outre, il est possible de supposer qu'au sein du projet il existe des affinités des agents deux à deux qui ont conduit à la formation du consortium. Enfin, la dernière raison qui plaide pour l'étude des collaborations deux à deux est la difficulté de modéliser la formation d'un projet, c'est-à-dire de tester la formation d'un sous-ensembles d'agents. Dans le cas, où l'échantillon contient N agents, il existe alors 2^N sous-ensemble d'agents. De plus, il faudrait pouvoir modéliser au sein du même modèle des consortiums de taille différente et aussi pouvoir prendre en compte le fait qu'un agent peut participer à plusieurs consortiums simultanément. Au final, l'hypothèse retenue est celle qui est traditionnellement retenue dans la littérature aussi bien dans les études statistiques (Breschi et Cusmano, 2004 ; Malerba et al., 2006 ; Fafchamps, van der Leij et Goyal, 2010 ; Breschi et Catalini, 2010 ; Deroian, M'Chirgui et Milleli, 2007 ; Roediger-Schluga et Barber, 2008) qu'économétriques (Autant-Bernard et al., 2007 ; Arranz et Fernandez de Arroyabe, 2008 ; Hanaki, Nakajima et Ogura, 2007 ; Paier et Scherngell, 2008).

D'autres études contournent le problème en ne s'intéressant pas aux partenaires des agents mais au fait qu'ils collaborent ou non quelles que soient les caractéristiques du (ou des) partenaire(s). Cassiman et Veugelers (2002), s'ils cherchent à expliquer quels sont les déterminants qui incitent une firme à co-opérer, adoptent une approche différente de celle adoptée traditionnellement dans les analyses réseaux. En effet, ils ne testent pas la formation d'un lien entre deux agents mais plutôt la probabilité qu'un agent possède au moins un accord de collaboration. Leur étude est basée sur les données Community Innovation Survey (CIS) de 1993. Les données qu'ils sélectionnent concernent les entreprises innovantes belges entre 1990 et 1992. Le principal apport de leur travail est d'estimer l'importance des externalités pour expliquer les déterminants des collaborations en R&D des entreprises. Les externalités sont divisés en deux grands types : "incoming spillovers" et "appropriability". Le premier effet renvoie à l'importance que les firmes accordent aux informations publiques disponibles telles que les informations sur les brevets, les publica-

tions et conférences de personnes spécialistes, les séminaires, ... Le second effet concerne la manière dont les firmes arrivent à s'approprier l'innovation. Pour cela, ils se basent sur les notes attribuées par les firmes dans le questionnaire de l'étude CIS concernant la capacité à protéger les innovations de produits et de procédés via 5 méthodes (brevets, secret, complexité, marques et copyright). Les auteurs montrent que ces deux types d'externalités ont un impact significatif et positif sur la probabilité de collaborer.

Les études qui portent sur les déterminants des choix de collaborations dans le cadre des PCRDT sont peu nombreuses. Autant-Bernard et al. (2007) se focalisent sur les décisions de collaboration des entreprises dans le domaine des nanotechnologies et établissent le rôle significatif de la structure du réseau de la période précédente et de la localisation spatiale des agents dans les décisions de collaboration. Paier et Scherngell (2008) étudient les déterminants des collaborations au sein des PCRDT à l'aide d'un modèle Logit. Ils cherchent à estimer le rôle joué par la proximité géographique, l'importance de connaître son partenaire, la proximité thématique, ... Ces études se distinguent sur plusieurs points. La première concerne la période étudiée. Paier et Scherngell travaillent sur la base sysres EUPRO Database qui au moment de leur étude comprenait l'ensemble des projets financés entre le 1^{er} PCRDT et le 5^{ème} PCRDT. Ils disposait de 70% des données concernant le 6^{ème} PCRDT, données qui ont été complétées par une étude menée sur un échantillon représentatif du 5^{ème} PCRDT par le centre de recherche autrichien de Vienne en 2007. La seconde différence concerne le choix de l'échantillon. Les auteurs ont sélectionné les agents parmi les répondants à l'étude menée par le centre de recherche de Vienne. A partir de ces agents, ils ont construits le réseau de collaborations dont ils ont extrait le 2-core (de Nooy, Mrvar et Batagelj, 2004) soit au final 191 agents. Les auteurs ont montré que l'effet réseau (pris en compte par l'indicateur de centralité de betweenness) ainsi que la proximité thématique et la proximité géographique affectaient de manière positive la probabilité de collaborer, l'effet le plus important étant dû à l'effet réseau.

Scherngell et Barber (2009) étudient les collaborations au sein des PCRDT mais ils se placent à un niveau plus agrégé celui des régions et cherchent à estimer quels sont les déterminants des collaborations entre régions. Ils retiennent comme variables indépendantes des caractéristiques propres à chaque région mais surtout ils cherchent à tester la pertinence de la proximité géographique et de la distance technologique. Ils montrent que la proximité géographique est un déterminant important des collaborations en R&D mais que l'effet le plus important est dû à la proximité technologique. Par la suite, les auteurs construisent le sous-réseau des collaborations entre entreprises et le

sous-réseau des collaborations entre universités. Ils montrent alors que l'effet proximité géographique est bien plus important dans le réseau des entreprises que dans celui des universités.

Dans cette partie, nous souhaitons évaluer l'influence des déterminants traditionnels ainsi que de l'effet réseau dans les choix de collaboration des agents au sein des PCRDT. L'accent sera mis sur la mesure des effets réseaux plus particulièrement afin de déterminer dans quelle mesure les collaborations passées d'un agent et son appartenance à un réseau affecte le choix des collaborations à la période présente. Pour montrer cela, nous nous intéressons uniquement aux agents qui ont participé aux 5^{ème} PCRDT et 6^{ème} PCRDT. Ces agents possèdent la caractéristique d'avoir déjà une expérience au sein des PCRDT. Nous cherchons à mettre en avant les déterminants des choix de collaborations uniquement pour ces agents. Afin de pouvoir répondre à cette question, il est nécessaire de posséder des données sur deux périodes distinctes pour connaître les collaborations passées d'un agent mais aussi la place qu'il occupe au sein du réseau. Nous disposons de ces données uniquement pour le sous-domaine des nanotechnologies qui est un sous-domaine de l'Electronique-Microélectronique. Cette restriction est nécessaire afin de posséder les données indispensables à l'estimation et à la mesure des différents effets. Même si l'échantillon est réduit, les résultats présentés par la suite pourront être généralisés à des échantillons plus importants en tenant compte de certaines spécificités. Par exemple, un élargissement de l'échantillon à l'ensemble des domaines technologiques et thématiques demanderait certainement de prendre en compte le domaine dans lequel se déroule la collaboration.

L'utilisation de l'économétrie n'a pas pour seule vocation de déterminer si les collaborations au sein des programmes cadres respectent ou non les déterminants traditionnels mais elle est aussi utile pour observer si ces déterminants s'appliquent de manière uniforme à tous les types d'agents ou plus exactement à tous les types de binôme puisque, dans cette partie, l'objet d'étude n'est plus l'agent individuel mais l'ensemble des binômes (paires d'agents) qu'il est possible de constituer à partir d'un échantillon donné d'agents. Ainsi, nous chercherons à montrer que selon les caractéristiques de ce binôme, la présence d'une collaboration antérieure et la composition du binôme en termes de type d'agents, les déterminants traditionnels peuvent affecter plus ou moins les choix des agents d'un binôme à collaborer.

L'existence ou non d'une collaboration entre deux agents est traditionnelle dans les études portant sur les réseaux de collaborations. Toutefois, s'en tenir à l'existence d'une collaboration ou non entre deux agents implique la perte d'une information importante. En effet, nous avons montré dans la pre-

mière partie que toutes les collaborations n'étaient pas aussi intenses les unes que les autres. Certains agents collaborent plus d'une fois ensemble. Aussi, il apparaît complémentaire de s'intéresser non pas à l'existence d'une collaboration entre deux agents mais à l'intensité de cette collaboration. Dans la thèse, l'objectif sera limité. En effet, nous ne cherchons pas à montrer qu'il existe des déterminants particuliers pour expliquer l'intensité des collaborations mais seulement à vérifier si les déterminants traditionnels de l'existence d'un lien bilatéral sont aussi susceptibles d'expliquer son intensité.

Étant donné le cadre particulier fourni par les collaborations au sein des PCRDT, il peut être intéressant de dépasser les relations bilatérales. Si l'on observe les conditions imposées par l'UE lors de la soumission d'un projet, on remarque qu'une d'entre elle revient sur de nombreux instruments, à savoir la présence d'au moins trois agents par projets issus de trois pays différents membres de l'UE ou en cours d'adhésion. La constitution d'un projet demande souvent la présence d'au moins trois agents. Il s'agit ici d'une condition minimal. Il est alors intéressant d'étudier non pas les déterminants des collaborations deux à deux mais les déterminants de la formation de triade ainsi que le respect des conditions imposées par l'UE.

Le réseau de collaborations des PCRDT est un réseau entièrement construit à partir d'une politique publique. Cela rend l'étude des déterminants des choix de collaborations intéressantes. La connaissance des déterminants permettra par la suite de mieux adapter les programmes cadres à ce que recherchent les agents y participant. De plus, il est possible de vérifier l'impact de certaines conditions imposées par l'UE, comme un minimum de trois agents membres de l'UE ou en cours d'adhésion dans chaque projet, sur les déterminants des choix de collaborations. Les éléments d'analyse réseau présentés en partie I associés à l'économétrie de réseaux permettront une telle approche.

La partie 2 de la thèse est organisée de la manière suivante. Le chapitre 1 vérifie si les déterminants rencontrés traditionnellement dans la littérature économique s'appliquent aussi aux PCRDT. Dans ce chapitre, nous présenterons les principales hypothèses que nous souhaitons tester. De plus, nous montrons que les déterminants des choix de collaborations peuvent affecter différemment les choix des binômes selon leurs caractéristiques propres. Le chapitre 2 s'intéresse non pas à l'existence ou non d'une collaboration entre deux agents mais à l'intensité de cette collaboration. Enfin, dans le chapitre 3, nous prenons le parti de sortir du cadre de l'analyse traditionnelle en nous intéressant, non pas à la formation de collaborations bilatérales, mais à la formation de triades.

Chapitre 3

Étude des déterminants des choix de collaborations bilatéraux

L'objectif est de déterminer quels sont les facteurs qui incitent deux agents à collaborer l'un avec l'autre au sein du réseau formé par les collaborations dans le sous-domaine des nanotechnologies dans le 6^{ème} PCRDT, réseau que nous avons étudié dans la partie précédente. Nous faisons l'hypothèse qu'un des facteurs explicatifs réside dans les collaborations passées des agents. Pour vérifier cette hypothèse, il faut posséder des données sur les réseaux précédents le réseau formé par le 6^{ème} PCRDT. C'est pourquoi nous utilisons le sous-domaine des nanotechnologies. C'est le seul domaine pour lequel nous possédons des données sur les deux périodes. La section 3.1 précise un certain nombre de points qui nous ont incité à faire ce choix. La section suivante présente les spécificités relatives aux réseaux lorsque l'on cherche à montrer quels sont les déterminants des collaborations entre deux agents. Puis, nous présenterons le principal modèle utilisé pour déterminer quels sont les facteurs susceptibles d'influencer les choix de collaborations d'un agent. Étant donné que nous nous intéressons à l'existence ou non d'un lien entre deux agents, nous utiliserons un modèle Logit. Enfin, la dernière section introduira l'ensemble des hypothèses que nous cherchons à tester dans nos modèles économétriques ainsi que les variables explicatives associées. En plus de l'hypothèse évoquée précédemment, nous chercherons à vérifier dans quelle mesure les choix de collaborations sont influencés par la proximité géographique.

3.1 Les données utilisées : prise en compte des effets réseaux grâce au 5^{ème} PCRDT

Jusqu'à présent, nous avons essentiellement étudié le réseau du 6^{ème} PCRDT. Dans la première partie, nous avons cherché à mettre en avant les caractéristiques de ce réseau et à déterminer quels étaient les agents centraux et les régions motrices. L'objectif de cette partie est de déterminer quelles sont les variables susceptibles d'influencer les agents dans le choix des partenaires dans le 6^{ème} PCRDT.

Des études (Fafchamps, van der Leij et Goyal, 2010 ; Autant-Bernard et al., 2007) montrent l'importance, pour expliquer les choix de collaboration à une période donnée, de connaître l'historique des collaborations entre ces agents aux périodes précédentes et, plus particulièrement, le fait que deux agents aient déjà collaboré à une période antérieure. Il est également possible de mesurer le rôle joué par la distance¹ entre deux agents dans le réseau formé par les collaborations d'une période antérieure sur les choix de collaboration à une période donnée (Hanaki, Nakajima et Ogura, 2007). Pour examiner si la place occupée dans les réseaux passés jouent un rôle dans les collaborations présentes, nous prenons comme référence le réseau formé par le 5^{ème} PCRDT (1998-2002).

Le choix du 5^{ème} et du 6^{ème} PCRDT² tient au fait qu'il existe une continuité aussi bien au niveau des domaines technologiques qu'au niveau des différentes définitions et notions utilisées. Par exemple, les différents critères qui définissent une Petite et Moyenne Entreprise (PME) sont identiques entre le 5^{ème} et le 6^{ème} PCRDT alors qu'ils diffèrent entre le 4^{ème} et le 5^{ème} PCRDT. Cette continuité est nécessaire afin d'éviter que l'étude soit biaisée. De plus, nous avons restreint notre analyse au domaine des nanotechnologies appliquées aux Technologies de la Société d'Information (TSI). Ce domaine, actuellement en pleine expansion, présente l'intérêt de toucher une grande diversité d'agents (entreprises, centres de recherche, ...). Nous avons retenu dans notre étude uniquement les agents qui ont participé à au moins un projet dans chaque programme cadre, soit 114 agents.

Les différentes régressions que nous réaliserons dans cette partie permettront d'étudier quels sont les déterminants qui favorisent les collaborations au sein des PCRDT pour un sous-ensemble d'agents qui sont ceux qui participent plusieurs fois aux PCRDT dans le même sous-domaine technologique.

¹Nous utilisons la notion de distance sociale, c'est-à-dire le nombre de liens minimum qui séparent deux agents dans un réseau.

²Nous remercions l'Association Nationale de la Recherche Technique (ANRT) pour nous avoir fourni les données ainsi que pour l'aide dans les traitements.

Ce sous-ensemble d'agents se distingue par une propension à collaborer plus importante comme le montre le tableau-ci-dessous.

	Nombre d'agents	Nombre de liens	Nombre de liens potentiels	Densité
FP5 Nanotechnologies	379	1619	71631	2,26%
FP6 Nanotechnologies	522	6591	135981	4,85%
Echantillon retenu	114	860	6441	13,35%

Tableau 34 : Nombre de liens par programme cadre dans le domaine des nanotechnologies.

Les 114 agents retenus forment un ensemble qui possède des propriétés proches de l'ensemble des agents participants au 6^{ème} PCRDT dans les nanotechnologies. Le tableau ci-dessous montre la répartition des agents en fonction de leur type.

	114 agents		FP6 Nanotechnologies	
Type d'agent	Nbre d'agents	Répartition	Nbre d'agents	Répartition
HE	41	36%	133	25%
IND	34	30%	118	23%
IND PME	12	11%	140	27%
OTH	5	4%	37	7%
RES	22	19%	94	18%
Total	114	100%	522	100%

Tableau 35 : Répartition des agents par type.

Dans l'échantillon retenu, les agents de l'enseignement supérieur ainsi que les entreprises sont sur-représentés à l'inverse des PME. Ceci explique notamment que les agents de l'échantillon aient une propension plus importante à collaborer. La proximité entre l'échantillon et l'ensemble des agents est plus marquée lorsqu'elle concerne la répartition géographique des agents comme le montre le tableau ci-dessous.

	Pourcentage d'agents	
	114 agents	FP6 Nanotechnologies
Union Européenne à 15 membres	89%	85%
Union Européenne à 27 membres	95%	91%
Cœur de l'UE	61%	62%

Tableau 36 : Répartition des agents selon leur localisation.

Sur ce tableau, on observe que la répartition des agents selon leur appartenance à un ensemble défini de pays comme l'Union Européenne est proche. De manière plus détaillée, on note juste une sous-représentation de l'Espagne parmi les pays du cœur de l'UE³. Sinon, derrière les 4 autres pays dominants en termes de participation, on retrouve les mêmes pays à savoir la Belgique, la Grèce, la Suisse, les Pays-Bas, l'Autriche et la Suède. Enfin, le dernier critère qui permet d'avancer une certaine proximité des 114 agents avec l'ensemble des agents ayant participé au 6^{ème} PCRDT dans les nanotechnologies correspond aux coûts supportés et aux financements reçus par les agents (en moyenne) par participation.

Moyenne	114 agents	FP6 Nanotechnologies
Coût supporté (en euros)	722453	710574
Financement reçu (en euros)	436872	414072

Tableau 37 : Coût et financement moyens par agent.

Sur le tableau, on observe que les deux sous-ensembles d'agents sont proches en termes de moyens engagés en R&D et de financement perçu.

Notons que, comme précisé en annexes, pour les grandes entreprises, nous avons considéré que la maison mère et ses filiales constituent autant d'agents différents contrairement à Hanaki, Nakajima et Ogura (2007). Cela permet notamment de tester des hypothèses liées à la proximité géographique. Ainsi, un agent peut choisir de collaborer avec une filiale d'un grand groupe parce qu'elle est géographiquement proche alors que ce même agent n'aurait noué aucun partenariat avec la maison mère qui peut être localisée à un endroit très différent. En outre, une filiale peut aussi posséder des compétences propres et/ou des procédés de fabrication qui peuvent influencer les choix des collaborations. Enfin, le regroupement sous un seul agent de la maison mère et de ses filiales pose un problème quand il s'agit d'un grand groupe dont les activités sont diverses et ne se situent pas forcément dans un même domaine technologique. En contrepartie, nous nous exposons à un biais qui résulte du fait que la prise de décision peut parfois être centralisée au sein de la maison mère.

Déterminer les facteurs influençant la probabilité que deux agents collaborent se heurte à une difficulté. Contrairement à un modèle Logit où la

³Le cœur de l'UE reprend la définition utilisée à la partie I à savoir les pays les plus participatifs. Le cœur est donc composé de l'Allemagne, l'Espagne, la France, l'Italie et le Royaume-Uni.

variable expliquée concerne un seul agent ou une seule entité, la variable expliquée dans notre cas concerne l'existence ou non d'une collaboration entre deux agents. L'entité étudiée n'est donc plus l'individu en lui-même mais une paire d'individus. La variable expliquée ainsi qu'une grande partie des variables explicatives renvoient donc au binôme formé par les deux individus. C'est la spécificité des modèles économétriques appliqués aux réseaux que nous présentons à la section suivante.

Malgré ces précautions, le modèle économétrique que nous testerons par la suite peut s'avérer moins adapté pour expliquer les collaborations de l'ensemble des agents participants au 6^{ème} PCRDT notamment car l'objectif de cette partie est de mettre en avant les déterminants des choix de collaborations pour un ensemble d'agents qui possède comme caractéristique particulière d'avoir déjà une expérience des PCRDT dans le domaine des nanotechnologies. Il faudra donc être prudent dans l'analyse des résultats.

3.2 Le modèle économétrique

3.2.1 Spécificités des modèles économétriques appliqués aux réseaux de collaborations.

Une difficulté importante dans la construction du modèle explicatif des choix de collaboration est que certaines variables explicatives que nous utilisons renvoient aux caractéristiques de chaque agent et non à la paire d'agents. Pour illustrer cette difficulté et les moyens d'y remédier, nous prenons comme exemple les variables correspondant aux indicateurs de centralité. Lorsque nous étudions le lien entre un agent i et un agent j dans le 6^{ème} PCRDT, nous possédons l'indicateur de centralité de l'agent i et celui de l'agent j lors du 5^{ème} PCRDT. Comme nous étudions la probabilité que deux agents collaborent, il nous est impossible d'introduire une variable pour l'agent i et une autre pour l'agent j . Cette difficulté est inhérente aux modèles économétriques qui portent sur des réseaux symétriques. Dans ce cas, il ne faut pas introduire de variables explicatives individuelles car l'indexation de ces variables compte dans l'estimation. En effet, une seule permutation peut modifier le résultat alors que le réseau étudié est identique. *“Here we encounter a practical difficulty that arises in all symmetric (undirectional) network regressions : since both authors occupy a symmetrical position in the coauthor pair, regressors must not depend on the order of indexation. This means that the same regressors must obtain if we reverse the order of i and j .”* (Fafchamps, van der Leij et Goyal, 2010, p23). Par conséquent, les régresseurs ne

doivent pas dépendre de l'ordre d'indexation au risque d'obtenir des estimations erronées. Nous devons construire les variables explicatives de manière à obtenir la même estimation quel que soit l'ordre d'indexation. Dans ce but, nous reprenons la méthode utilisée par Fafchamps, van der Leij et Goyal (2010).

A partir des données individuelles sur les agents, nous construisons deux types de variables :

1. La valeur absolue de la différence entre la valeur des variables des deux partenaires i et j C_i et C_j (l'important ici étant l'écart entre les valeurs, le recours à la valeur absolue évite tout problème lié au signe) :

$$\Delta C = |C_i - C_j|$$

Si le coefficient du paramètre associé à ΔC dans l'estimation est positif alors plus les agents seront différents en termes de centralité plus la probabilité qu'ils collaborent sera forte. C'est un des résultats obtenus par Goyal, van der Leij et Moraga-Gonzalez (2004) dans le cas des réseaux de co-auteurs dans le domaine scientifique. Par exemple, ce résultat peut s'expliquer, dans le domaine des publications scientifiques, par l'importance des relations entre les doctorants et les directeurs de thèse. A l'inverse, si le coefficient est négatif, alors plus les agents seront différents en termes de centralité, moins la probabilité qu'ils collaborent sera forte.

2. La deuxième variable est égale à la moyenne des valeurs de chaque agent :

$$\overline{C} = \frac{C_i + C_j}{2}$$

Si le signe du paramètre associé à \overline{C} dans l'estimation est positif alors deux agents avec de forts indicateurs de centralité sont attractifs l'un pour l'autre. Au contraire, si le signe du paramètre associé à \overline{C} est négatif alors des agents centraux n'ont pas intérêt à collaborer. Nous parlons ici des paires d'agents centraux car ce sont les agents pour lesquels l'impact de \overline{C} sur la probabilité de collaborer est le plus significatif.

Grâce à ces deux types de variables, l'ordre d'indexation n'intervient plus dans le modèle. Les variables explicatives que nous utilisons sont pour la plupart construites de cette manière. Dans la section suivante, nous présentons

l'ensemble des variables dont nous allons estimer les effets par la suite et, surtout, les hypothèses associées à chacune d'elles.

3.2.2 Quel modèle utiliser pour déterminer les facteurs influençant les choix de collaborations dans les PCRDT ?

Modèle logit et variable latente

Nous cherchons à expliquer quels sont les facteurs qui influent sur la collaboration entre deux agents. La variable explicative de notre modèle, notée y_{ij} , est une variable binaire qui prend la valeur 1 si les agents i et j collaborent dans le 6^{ème} PCRDT et 0 sinon. L'ensemble des collaborations est représenté dans le graphe E dans l'annexes A.13.

La variable expliquée ne fait pas référence à un seul agent mais renvoie à un binôme constitué par deux agents. La taille de l'échantillon n'est donc pas égale au nombre d'agents (114) mais au nombre de paires d'agents que l'on peut former, soit 6441 binômes potentiels. Comme nous l'avons précisé à la section précédente, dans notre cas, il est équivalent d'utiliser y_{ij} ou y_{ji} puisque nos variables, dépendantes ou explicatives ne dépendent pas de l'ordre d'indexation.

$$y_{ij} = \begin{cases} 1 & \text{si l'agent } i \text{ et l'agent } j \text{ collaborent, } i \neq j \\ 0 & \text{sinon} \end{cases}$$

Sous forme matricielle, le modèle s'écrit :

$$Y = X\beta + \epsilon$$

avec Y la matrice des variables dépendantes, X la matrice des variables explicatives, β le vecteur de paramètres et ϵ le vecteur des perturbations. Dans la sous-section suivante, nous détaillerons la construction de la matrice X . On a $y_{ij} = X_{ij}\beta + \epsilon_{ij}$.

La variable explicative étant binaire, on ne peut donc pas utiliser la méthode des Moindres Carrés Ordinaires (MCO). En effet, comme la variable explicative prend la valeur 0 ou 1, le terme d'erreur est égale soit à $1 - X_{ij}\beta$ soit à $-X_{ij}\beta$. Les résidus ne sont donc pas normalement distribués. Il y a donc un problème dans l'application des tests usuels de significativité qui reposent sur l'hypothèse de normalité des perturbations (test de Student et test de Fisher notamment).

De plus, ce modèle conduit à une violation de l'hypothèse d'homoscédasticité des perturbations ($Var(\epsilon_{ij}) = \sigma_\epsilon^2 I_N, \forall ij$). La variance de la perturbation n'est pas identiquement distribuée. On a :

$$\begin{aligned} Var(\epsilon_{ij}) &= X_{ij}\beta(1 - X_{ij}\beta)^2 + (1 - X_{ij}\beta)(-X_{ij}\beta)^2 \\ &= X_{ij}\beta(1 - X_{ij}\beta) \\ &= E(y_{ij})(1 - E(y_{ij})) \end{aligned}$$

La variance n'est pas identique pour tout couple ij puisqu'elle dépend de y_{ij} . Une solution consiste à considérer que la régression est de la forme :

$$y_{ik}^* = X_{ij}\beta + \epsilon_{ij}$$

où y_{ik}^* est le profit escompté par l'agent i de sa collaboration avec j . Le problème est que cette variable n'est pas observée. On observe s'il y a collaboration ou non entre deux agents. La variable y_{ij}^* est alors appelée variable latente. C'est cette variable qui est censée dépendre linéairement des variables explicatives. On fait alors l'hypothèse qu'une firme i collabore avec une firme j si et seulement si le profit qu'elle retire de cette collaboration est supérieur à un certain seuil que l'on peut normaliser à 0. On a alors :

$$\begin{cases} y_{ij} = 1 & \text{si } y_{ik}^* > 0 \\ y_{ij} = 0 & \text{si } y_{ik}^* \leq 0 \end{cases}$$

Pour estimer les paramètres en jeu dans cette régression, on utilisera un modèle Logit. On suppose donc que les perturbations suivent une loi logistique. Le modèle Logit va s'intéresser à la probabilité que deux firmes i et j collaborent, c'est-à-dire la probabilité que $y_{ij} = 1$ si et seulement si $Prob(\pi_{ij} \geq 0)$. Le modèle ne détermine pas exactement la réalisation de l'événement $y_{ij} = 1$ ou $y_{ij} = 0$ mais fournit plutôt une mesure théorique de la proportion d'observations pour lesquelles cet événement s'est réalisé. Dans un modèle Logit, la variance des perturbations est normalisée à $\pi^2/3$. Du fait de cette normalisation, seuls les signes des paramètres des différentes variables étudiées seront réellement pertinents. On peut tout de même extraire les effets marginaux mais ils ne sont pas donnés directement par les coefficients. Pour obtenir la valeur d'un effet marginal, on dérive la probabilité estimée par rapport aux composantes de X .

La fonction de densité d'un modèle Logit est la suivante :

$$\lambda(X_{ij}\beta) = \frac{\exp(X_{ij}\beta)}{(1 + \exp(X_{ij}\beta))^2}$$

La fonction de répartition du modèle Logit est :

$$\Lambda(X_{ij}\beta) = \frac{\exp(X_{ij}\beta)}{1 + \exp(X_{ij}\beta)}$$

On cherche le vecteur β qui permet de maximiser la log-vraisemblance du modèle. La vraisemblance du modèle est la fonction de densité conjointe des observations. Ces dernières étant indépendantes les unes des autres, la vraisemblance s'écrit comme le produit des fonctions de densité de chacune des observations. On obtient :

$$L(Y; X, \beta) = \prod_{ij} \left(\frac{1}{1 + \exp(X_{ij}\beta)} \right)^{1-y_{ij}} \left(\frac{\exp(X_{ij}\beta)}{1 + \exp(X_{ij}\beta)} \right)^{y_{ij}}$$

De cette expression, on peut déduire la log-vraisemblance :

$$\log L(Y; X, \beta) = \sum_{ij} (\log(1 + \exp(X_{ij}\beta)) + y_{ij}(X_{ij}\beta))$$

On cherche alors le vecteur de paramètre β qui annule la dérivée de la fonction de log-vraisemblance. Cependant dans le cas du modèle Logit, il n'y a pas de solution analytique. On procède donc par itération.

Interprétations des estimations et effets marginaux

Dans les modèles Logit, comme dans tous les modèles non linéaires, les valeurs des paramètres ne sont pas directement interprétables, seuls les signes des paramètres estimés sont utilisables sans aucun traitement. Si la valeur du paramètre est positive, cela signifie que la variable associée à ce paramètre influence de manière positive la probabilité que deux agents collaborent et inversement en cas de paramètre négatif. Pour déterminer plus précisément l'effet d'une variable, il faut calculer les effets marginaux. Les effets marginaux s'obtiennent de la manière suivante :

$$\frac{\partial E[y|X]}{\partial X} = \Lambda(X_{ij}\beta)[1 - \Lambda(X_{ij}\beta)]\beta$$

Pour calculer ces effets marginaux, il existe deux possibilités (Greene, 2005). La première possibilité consiste à évaluer ces effets marginaux aux valeurs moyennes des données. La seconde possibilité consiste à évaluer les effets marginaux pour chaque observation et ensuite calculer la moyenne des effets marginaux individuels pour l'échantillon. Pour les échantillons de taille importante, les deux méthodes donnent des résultats identiques mais lorsque

l'échantillon est de faible taille, il est préférable de calculer la moyenne des effets marginaux individuels.

Enfin, dans le cas où les régresseurs sont des variables muettes, il faut alors calculer les effets marginaux d'une manière différente (même si la méthode ci-dessus fournit une première approximation). Ceci s'explique par le fait que la dérivée est calculée pour de petites variations qui ne conviennent pas aux variables muettes qui ne peuvent par définition ne prendre que deux valeurs à savoir 1 et 0. Pour une variable muette m , l'effet marginal se calcule de la manière suivante (Greene, 2005) :

$$\text{Effet marginal} = P[Y = 1 | \bar{X}_{(m)}, m = 1] - P[Y = 1 | \bar{X}_{(m)}, m = 0]$$

où $\bar{X}_{(m)}$ représente la moyenne des autres variables explicatives du modèle.

Dans la section suivante, nous mettons en avant les différentes hypothèses (et variables associées) que nous souhaitons tester dans notre modèle ainsi que les estimations associées.

3.3 Le rôle primordial des réseaux sociaux au sein des PCRD et l'effet mitigé de la proximité géographique

3.3.1 Formulation des hypothèses testables : des réseaux sociaux à la proximité géographique

Dans notre modèle, la collaboration entre deux agents dépend de trois types de variables : les variables réseaux, les variables liées aux moyens en R&D et les variables géographiques. L'introduction de chacune de ces variables dans le modèle concernant les différents déterminants des collaborations vise à vérifier la pertinence des hypothèses formulées.

Les variables réseaux

On peut supposer que les collaborations à une période donnée dépendent de la position des agents au sein du réseau de collaboration de la période précédente. Nous mesurons la position d'un agent au sein d'un réseau par deux types de variables, chacune associée à une hypothèse testable.

La première hypothèse que l'on souhaite vérifier est simplement le fait qu'un agent central est un partenaire attractif pour les autres agents en termes de collaborations. Nous utilisons des indicateurs de centralité qui sont propres à chaque agent.

Hypothèse 1a : La probabilité que deux agents collaborent au cours du 6^{ème} PCRDT est fonction de la centralité des agents au sein du réseau formé par le 5^{ème} PCRDT.

Afin de tester cette hypothèse, nous utilisons une mesure de centralité qui est le degré renforcé⁴ (Billand, Frachisse et Massard, 2008). Il est impossible d'utiliser l'"eigenvector centrality" dans la mesure où le réseau associé au 5^{ème} PCRDT n'est pas connexe. Dans la partie 1, nous avons utilisé l'"eigenvector centrality" car la grande composante contenait quasiment tous les agents du réseau. Parmi les 114 agents, certains n'appartiennent pas à cette composante et nous aurions alors aucun indicateur pour ces agents. Si au moins un agent n'appartient pas à la composante principale, il est impossible de calculer la variable. Pour contourner cette difficulté, nous utilisons le degré renforcé.

Un agent central selon l'indicateur de degré renforcé est un agent qui possède de nombreux liens avec des agents qui sont eux mêmes centraux. On peut penser qu'un tel agent jouit d'un certain prestige et d'une certaine influence qui peuvent être nécessaires à la bonne réalisation de la collaboration. De plus, il est intéressant d'examiner la nature des relations entre agents centraux. Comme cette mesure est propre à chaque agent, nous construisons deux variables distinctes à partir de la méthode décrite dans la section 3.2 : \overline{DR} et ΔDR .

La variable \overline{DR} renvoie notamment à la notion de "preferential attachment" développée par Barabasi et Albert (1999). Les agents avec un nombre important de collaborations à une période donnée ont une probabilité plus élevée d'attirer plus de partenaires à la période suivante. Cet effet peut s'ex-

⁴Le degré renforcé (Frachisse, 2007) est basé à la fois sur le degré des sommets et sur la distance géodésique entre tous les sommets du réseau. L'objectif est d'obtenir un indicateur de centralité qui prenne en compte à la fois le degré du sommet en question et celui des sommets auxquels il est lié de manière directe et indirecte. La centralité d'un agent dépend de celle de tous les autres agents de manière inversement proportionnelle à la distance qui les sépare. Le graphe J en annexes A.13 représente le réseau de collaboration du 6^{ème} PCRDT en prenant en compte la centralité des agents.

Cette mesure, dans sa conception, se rapproche de l'indicateur de centralité obtenu par la méthode des valeurs propres (Bonacich, 1972a, 1972,b) et de l'indicateur de "power index" (Bonacich, 1987) qui utilise la centralité des plus proches voisins pour calculer la centralité d'un agent.

pliquer par un phénomène de reconnaissance qui s'établit dans le réseau, l'agent central à une période donnée devenant alors un acteur important du réseau reconnu par les autres agents. De cette manière, on peut voir apparaître des collaborations entre des agents centraux et des agents plus périphériques mais aussi des collaborations entre agents centraux. Puisque l'on calcule la moyenne pour deux agents de l'indicateur de centralité, l'effet de cette variable sur la probabilité de collaborer sera d'autant plus fort que les indicateurs de centralité des deux agents seront élevés. La variable ΔDR indique si deux agents qui collaborent sont plutôt semblables ou non en termes de centralité. Cette variable renvoie à un phénomène observé dans les réseaux de co-auteurs dans lesquels les agents centraux qui co-écrivent beaucoup sont souvent liés avec des agents qui possèdent moins de co-auteurs. Dans le cas des publications scientifiques, on peut citer les nombreuses co-publications entre doctorants et directeurs de thèse.

Les variables \overline{DR} et ΔDR sont nécessaires pour comprendre le rôle que joue le réseau formé par le 5^{ème} PCRDT dans les choix de collaborations du 6^{ème} PCRDT, notamment sur le prestige et l'attractivité que peut octroyer une position centrale au sein d'un réseau.

La deuxième hypothèse prend en compte le fait que le réseau formé par le 5^{ème} PCRDT permet aux agents par le biais de leur collaboration passée d'obtenir des informations sur d'autres agents. Nous cherchons à déterminer dans quelle mesure la proximité de deux agents dans le réseau antérieur (la proximité sociale) influence leur décision de collaborer ensemble dans le réseau du 6^{ème} PCRDT. Cette proximité sociale peut se caractériser de plusieurs manières comme une collaboration entre les deux agents à la période précédente ou encore l'existence de partenaires communs dans le réseau du 5^{ème} PCRDT.

Hypothèse 1b : La présence à une période antérieure d'un lien entre deux agents et/ou la proximité au sein de ce réseau entre deux agents favorise(nt) les collaborations futures.

Pour illustrer cette hypothèse, nous représentons le graphe du 5^{ème} PCRDT (cf. graphe F en annexes A.13). De plus, nous avons construit un graphe représentant uniquement les collaborations qui existent à la fois dans le 5^{ème} PCRDT et dans le 6^{ème} PCRDT (graphe G en annexes A.13). Nous replaçons ces collaborations au sein du graphe des collaborations du 5^{ème} PCRDT (graphe I annexes A.13) et du 6^{ème} PCRDT (graphe H annexes A.13).

Nous utilisons plusieurs variables différentes pour tester cette hypothèse.

La première variable *LIEN5* sert à tester le fait que si deux agents collaborent à une période, alors la probabilité qu'ils collaborent à nouveau à la période suivante s'accroît toute chose étant égale par ailleurs. Ce prolongement de la collaboration peut s'expliquer par la confiance mutuelle des deux agents l'un envers l'autre. Un agent i peut avoir confiance en un agent j parce qu'il est assuré de la bonne volonté et du sérieux de l'agent j dans le travail et/ou alors il pense que l'agent j possède les compétences suffisantes (Bertacchini, 2000 ; Suire et Al., 2006). La confiance entre deux agents qui s'établit lors d'une collaboration est susceptible de les inciter à poursuivre cette collaboration à la période suivante. Cette variable est construite de la manière suivante :

$$\begin{cases} LIEN5_{ij} = 1 & \text{si } i \text{ et } j \text{ possèdent un lien dans le 5}^{\text{ème}} \text{ PCRDT} \\ LIEN5_{ij} = 0 & \text{sinon} \end{cases}$$

La deuxième variable *PARTCOM* correspond au nombre de partenaires communs à deux agents dans le réseau formé par le 5^{ème} PCRDT. Cette variable est construite de la manière suivante :

$$\begin{cases} PARTCOM_{ij} = 1 & \text{si } i \text{ et } j \text{ possèdent au moins un partenaire} \\ & \text{commun dans le 5}^{\text{ème}} \text{ PCRDT} \\ PARTCOM_{ij} = 0 & \text{sinon} \end{cases}$$

Quel est l'intérêt de cette variable ? Supposons qu'à une période donnée, il existe trois agents i , j et k tels que les deux premiers agents ne collaborent pas entre eux mais collaborent avec k dans deux projets distincts. Les collaborations de l'agent k sont l'occasion pour ce dernier de recueillir des informations sur l'agent i (respectivement sur l'agent j) dont il pourra faire part à l'agent j (respectivement à l'agent i). Ces informations pourront alors être utilisées par l'agent i et par l'agent j dans le choix de leurs partenaires à la période suivante. L'impact du réseau de connaissances a été mis en évidence par Gulati (1993, 1998) lors d'interviews de responsables exécutifs chargés des partenariats. L'un des responsables a déclaré : “*In some cases we realize that perhaps our skills don't really match for a project, and our partner may refer us to another firm about whom we were unaware... An important aspect of this referral business is of course about vouching for the reliability of that firm. Thus, if one of our long standing partners suggests one of their own partners as a good fit for our needs, we usually consider it very seriously.*” (1993, p84). Cette citation montre le rôle important joué par les partenaires d'un agent dans ses choix de collaboration.

Cette dernière variable, bien que pertinente pour tester l'hypothèse, ne rend compte que partiellement de l'effet réseau. En effet, pour obtenir des informations sur un partenaire potentiel, un agent mobilise ses contacts. Si ces derniers ne possèdent pas directement de l'information sur le partenaire potentiel alors peut-on considérer pour autant que l'effet réseau est nul ? Le réseau doit être perçu comme une chaîne qui relie les agents les uns aux autres. De fait, si les contacts d'un agent ne possèdent pas directement des informations sur un partenaire potentiel, on peut imaginer que ces derniers vont à leur tour mobiliser leurs propres contacts et ainsi de suite jusqu'à l'obtention d'une information. Bien sûr, comme dans la partie précédente, il est tout à fait réaliste de penser qu'il existe une limite ou une perte de pertinence qui est croissante avec le nombre de liens mobilisés pour obtenir l'information. Plus la chaîne qui relie deux agents est longue plus la probabilité d'obtenir des informations (de qualité) est faible du fait qu'à chaque lien mobilisé il peut exister des frictions qui empêchent l'information de circuler. Afin de prendre en compte l'ensemble de ces remarques dans notre estimation, il est possible d'utiliser la distance géodésique ($DG5$) qui sépare deux agents dans le 5^{ème} PCRDT. Ainsi, on suppose que l'effet réseau sera d'autant plus fort que deux agents seront situés à proximité dans le réseau de la période précédente. Afin d'éviter des problèmes lors de l'estimation dans le cas où les deux agents n'appartiennent pas à la même composante dans le réseau, nous prenons, non pas la distance géodésique, mais l'inverse de cette distance. Comme dans la partie précédente, nous supposons que si deux agents n'appartiennent pas à la même composante, la distance entre les deux est égale à l'infini.

Remarque : Les trois variables ci-dessus ne peuvent être introduites simultanément même deux à deux. En effet, un lien entre deux agents à la période précédente implique nécessairement une distance géodésique de 1 et, de même, la présence d'un partenaire commun à la période précédente implique une distance géodésique de 1 ou 2 à cette même période.

Les deux premières hypothèses servent à vérifier l'importance du réseau social des agents dans les choix de collaboration. Dans la sous-section suivante, nous présentons les variables associées aux moyens en R&D.

Les variables liées aux moyens en R&D

On suppose que la probabilité que deux agents collaborent est liée aux moyens que sont prêts à consacrer à la collaboration. Les agents, au moment de choisir leurs partenaires à la date t , possèdent des informations (comptes

publiés, brochures,...) sur les années antérieures. Les agents basent alors leurs choix à la période t sur les éléments en leur possession, par exemple les dépenses en R&D antérieures, qui donnent une idée assez précise des moyens dont peuvent disposer les partenaires potentiels à la période t . Les agents peuvent alors utiliser cette information afin de réaliser leurs anticipations sur les moyens mis en œuvre par les autres agents dans le cadre du 6^{ème} PCRDT. On peut qualifier ces anticipations d'extrapolatives.

Hypothèse 2 : La probabilité que deux agents collaborent est fonction des moyens dont chacun dispose dans ses activités de R&D.

Dans notre base de données, nous ne disposons pas de variables indiquant le nombre de chercheurs ou les dépenses annuelles en R&D de chaque agent, variables qui auraient pu être pertinentes. C'est pourquoi, nous utilisons les coûts supportés par les agents dans le cadre du 5^{ème} PCRDT comme proxy des moyens qui sont mis en œuvre. Le graphe K (annexes A.13) illustre le réseau des collaborations du 6^{ème} PCRDT en représentant les agents en fonction des moyens dont ils disposent.

Nous supposons que plus un agent supporte un coût élevé dans le cadre du PCRDT, plus il dispose de moyens en R&D. Ainsi, nous avons construit les variables \overline{COUT} et $\Delta COUT$. Nous utilisons les données relatives aux coûts supportés plutôt que celles relatives au financement demandé car s'il est possible, dans les deux cas, qu'au sein de chaque projet la règle d'équité-répartition soit utilisée, il est plus probable que celle-ci ne soit pas utilisée pour le coût étant donné les différences de type entre les agents.

La variable est calculée comme la moyenne des coûts supportés par les agents par participation. Nous aurions pu prendre la somme mais cette dernière aurait introduit plusieurs biais. Le premier est que le coût total aurait pu être fortement corrélé avec les indicateurs réseaux dans la mesure où plus un agent participe plus il supporte un coût important. La seconde est que la somme n'aurait reflété qu'imparfaitement les moyens en R&D dont dispose l'agent puisque certains agents peuvent réaliser une part de leur R&D en dehors des PCRDT. Dans la première partie, nous avons montré que des agents avaient augmenté leur participation de manière significative entre les deux programmes cadres. A l'inverse d'autres ont baissé significativement leur participation. De fait, le coût total supporté aurait été une mesure trop imparfaite pour rendre compte de la réalité du 6^{ème} PCRDT.

En revanche, la moyenne des coûts supportés par participation présente l'avantage d'être normalisée puisqu'elle prend en compte le nombre de participations des agents. Ainsi, elle donne une mesure correcte de ce qu'un agent

est capable de supporter comme coût dans une participation, ce qui est finalement l'information que souhaite obtenir les partenaires potentiels. Certes, on peut avancer l'argument que si un agent participe à plusieurs projets dans le 5^{ème} PCRDT alors c'est aussi le signe qu'il possède des moyens en R&D importants mais cette capacité à participer à de nombreux projets est indirectement prise en compte par les indicateurs réseaux qui sont introduits dans le modèle et qui renvoient au prestige acquis par un agent dans le 5^{ème} PCRDT.

Il est possible d'utiliser le coefficient de rang de Spearman (1904) pour montrer que si l'on classe les agents en fonction du coût moyen par participation puis en fonction du coût total qu'ils supportent alors le rang occupé par les agents est proche. En effet, le coefficient de rang de Spearman est de 0,83. Or, ce coefficient varie entre 0 et 1. Plus il est proche de 1 plus cela signifie que les deux classements sont liés l'un à l'autre. Dans notre cas, cela signifie que les rangs occupés par les agents selon ces deux variables sont liés. De ce fait, le classement des agents selon les deux critères est proche. Les agents qui possèdent un coût moyen par participation élevé sont ceux qui possèdent un coût total élevé.

Nous n'utilisons pas les données concernant les coûts supportés dans le 6^{ème} PCRDT et ce pour plusieurs raisons. Premièrement, un agent peut avoir des difficultés à déterminer les moyens dont disposent un partenaire potentiel à un l'instant t . Il est souvent plus facile d'obtenir des données sur les périodes passées que sur la période présente via différentes bases de données ou les informations publiées sur les sites internet. Deuxièmement, le 6^{ème} PCRDT s'étale sur 5 ans. Certaines collaborations peuvent avoir lieu en début de période alors que le coût moyen est calculé sur l'ensemble des 5 ans. Ce décalage pourrait introduire un biais dans certains cas, notamment si les moyens en R&D d'un agent augmentent au fil des années. Dans ce cas il serait plus élevé qu'il ne l'était au moment où la collaboration a été réalisée et ne reflèterait alors pas la réalité au moment où la collaboration a été effectuée. Troisièmement, si on prenait en compte les coûts supportés dans le 6^{ème} PCRDT, si deux agents collaborent et donc participent au même projet alors il est possible que le coût qu'ils supportent au sein du projet soit identique, notamment si ce dernier est partagé en parts égales ce qui biaiserait l'effet relatif aux variables \overline{COUT} et $\Delta COUT$. En prenant en compte les coûts supportés dans le 6^{ème} PCRDT, on risquerait ainsi d'avoir des agents aux profils identiques en termes de moyens mis en œuvre en R&D.

Ces variables sont donc calculées sur le 5^{ème} PCRDT. De cette manière, nous pouvons ainsi observer dans le 6^{ème} PCRDT si les agents collaborent de façon préférentielle avec des agents qui ont supporté un coût élevé lors du

programme cadre précédent. Dans ce cas, les agents chercheraient à s'associer avec des partenaires qui auraient des moyens importants. En retardant cette variable d'une période, on pourrait observer la manière dont se forment les binômes.

Ce type de raisonnement est courant dans les papiers qui s'intéressent aux collaborations entre agents quel que soit le type de la collaboration (en recherche et développement, co-auteurs,...). Les auteurs cherchent à identifier si les agents collaborent avec des agents semblables ("assortativité matching") ou au contraire recherchent dans la collaboration un agent aux caractéristiques différentes. Les types de caractéristiques mises en avant sont différents. Ainsi dans les réseaux de co-auteurs, notamment en économie, la différenciation se fait sur la productivité des auteurs de manière à essayer d'identifier des schémas où des auteurs réputés travaillent entre eux mais aussi avec des auteurs moins qualifiés. Certains travaux mettent en avant l'argument que l'auteur le moins reconnu fournit une part de travail plus importante en échange de la qualité du travail de l'auteur le plus connu (Goyal, van der Leij et Moraga-Gonzalez, 2006).

La différenciation entre les auteurs se fait aussi en fonction des champs de recherche de ces derniers (Fafchamps, van der Leij et Goyal, 2010). Dans le cadre de collaborations en R&D, la différenciation peut porter sur le secteur dans lequel travaillent les agents. Dans les deux cas (réseaux de co-auteurs ou réseaux de collaborations), les agents peuvent être amenés à collaborer avec des agents qui travaillent dans des secteurs, champs ou disciplines complémentaires des leurs afin de répondre à l'exigence de la recherche actuelle qui se veut de plus en plus pluri-disciplinaire. Dans notre cas, il aurait été intéressant de tester une telle hypothèse en croisant la base PCRDT avec une base de données susceptible de nous renseigner sur les secteurs d'activités des différents agents comme la base de données Diane qui donne le secteur d'activité (Code NAF (Nomenclature d'Activités Française)) pour les établissements principaux privés français par exemple. Toutefois, une telle démarche se heurte à plusieurs difficultés : la difficulté d'identifier les agents (cf. annexes 3) et la difficulté de relier les bases du fait de l'absence d'identifiant du type SIRET ou SIREN pour les entreprises françaises par exemple. En effet, l'absence d'un tel identifiant au niveau européen et l'absence d'homogénéité au niveau européen en termes de nomenclatures d'activité rendent ce travail de correspondance ardu.

Nous venons de présenter deux des déterminants principaux identifiés dans la littérature. Toutefois, un autre élément est susceptible d'influencer les choix de collaborations des agents, il s'agit de la proximité géographique.

les hypothèses relatives à ce facteur explicatif potentiel des choix de collaborations sont présentées dans la sous-section suivante.

Les variables géographiques

La distance géographique entre deux agents est susceptible d'influencer la probabilité de collaborer. De nombreuses études ont ainsi montré l'impact de la proximité géographique sur l'innovation et les interactions entre agents. Cette proximité permet de faciliter les échanges notamment les contacts directs entre agents (Zucker and al., 1994 ; Almeida and Kogut, 1999 ; Singh, 2005 ; Breschi and Lissoni, 2009). De plus, une telle proximité permet de réduire les coûts (coûts de transport et de communication) (Englmann et Walz, 1995 ; Martin et Ottaviano, 1999 ; Baranes et Tropeano, 2003) et de réaliser des gains d'efficacité dans la collaboration (Colletis et Winterhalter, 1991 ; Graitson, 2000). Dans un cadre comme celui des PCRDT qui cherchent à favoriser l'innovation afin *in fine* d'améliorer la compétitivité des firmes européennes, les effets bénéfiques de la proximité géographique sur les coûts et l'efficacité économique peuvent inciter les agents lorsqu'ils choisissent leurs partenaires à prendre en compte la distance qui les sépare. Les collaborations entre agents peuvent alors être favorisées par la proximité géographique puisque celle-ci peut permettre d'améliorer l'efficacité de la collaboration. On peut alors supposer qu'un agent français collaborera plus aisément avec un autre agent français plutôt qu'avec un agent étranger. De même, un agent français collaborera plus avec un agent allemand qu'avec un agent suédois étant donné la proximité géographique entre les deux pays.

Dans le cas des PCRDT, il faut toutefois se montrer prudent car les programmes cadres présentent pour spécificité de posséder un certain nombre de critères visant à favoriser les collaborations entre agents de pays différents. Dans ce contexte, la proximité géographique n'est pas forcément un facteur déterminant pour le choix des partenaires, ses impacts peuvent être plus délicats à appréhender.

Hypothèse 3 : La probabilité que deux agents collaborent est d'autant plus forte qu'ils sont géographiquement proches l'un de l'autre.

Pour tester cette hypothèse, nous avons construit plusieurs variables.

La variable *PROXGEO* est la principale variable utilisée pour vérifier si la proximité géographique influe ou non sur la probabilité de collaborer. Si les deux agents sont de nationalité identique ou issus de pays frontaliers alors la variable prend la valeur 1 et 0 sinon.

$$\begin{cases} PROXGEO_{ij} = & 1 \text{ si } i \text{ et } j \text{ sont localisés dans le même pays} \\ & \text{ou dans des pays frontaliers} \\ PROXGEO_{ij} = & 0 \text{ sinon} \end{cases}$$

Le fait que cette variable prenne en compte la possibilité que deux agents appartiennent au même pays peut paraître surprenant dans le cas des PCRDT puisque ces programmes ont pour vocation de favoriser les coopérations entre agents de pays différents notamment en mettant en avant des conditions qui obligent chaque projet à compter au moins trois agents de nationalités différentes.

Le nombre moyen de partenaires par projet étant bien supérieur à trois, il n'est pas impossible de voir apparaître, au sein des projets, des collaborations entre agents de même nationalité. Cette collaboration peut même constituer la base du projet, les partenaires additionnels étant choisis alors selon des critères permettant le financement du projet. Une manière de vérifier cette hypothèse est de construire une variable (*IDPAYS*) qui prend la valeur 1 si et seulement si les deux agents appartiennent au même pays. Dans la mesure où nous nous intéressons seulement aux agents présents dans les deux programmes cadres, si une telle variable a un effet positif et significatif, cela signifie que le dispositif de l'UE favorise les collaborations entre agents de pays différents peut être contourné en créant des projets d'envergure suffisante pour satisfaire aux conditions de l'UE.

Enfin, pour compléter l'analyse, nous construisons la variable *BORDER* qui prend la valeur 1 si les deux agents appartiennent à des pays frontaliers et 0 sinon. Si le paramètre associé à cette variable est positif et significatif et que dans le même temps la variable *IDPAYS* est non significative alors cela signifiera que les contraintes introduites par l'UE guident réellement les choix des partenaires. Dans le cas contraire, cela signifie que les agents arrivent à contourner cette contrainte.

Remarque : Les variables *BORDER* et *IDPAYS* ne peuvent être introduites en même temps que la variable *PROXGEO* car si une des deux premières variables citées prend la valeur de 1 pour un binôme (i, j) alors la variable *PROXGEO* sera également égale à 1.

Toutefois, ces variables peuvent s'avérer insuffisantes pour mesurer la proximité géographique. Certains pays proches les uns des autres ne possèdent pas pour autant de frontières communes. L'effet de proximité peut alors jouer sans qu'il soit détecté par la variable *PROXGEO*. Afin de combler cette lacune, nous créons deux variables pour tenter de mesurer l'effet

de la proximité géographique *DISTANCE* et *TEMPS*. La première renvoie à la distance exprimée en kilomètres entre les deux agents du binôme alors que la seconde est une évaluation du temps exprimée en minutes qu'il faut pour parcourir cette distance. Dans les deux cas, nous avons utilisé des trajets par la route pour estimer la distance et le temps entre les agents. Ces deux variables renvoient donc à une approximation assez imparfaite. Il aurait été préférable de calculer pour chaque binôme le temps par le moyen de transport le plus avantageux.

Remarque : Pour déterminer la distance entre deux agents, nous avons calculé la distance entre les villes principales des NUTS2 dans lesquelles ces agents sont localisés. Par exemple, si un agent est localisé à Grenoble et un autre à Londres, nous avons calculé la distance entre les villes les plus importantes des régions auxquelles appartiennent les deux agents. Dans cet exemple, Grenoble appartient à la région Rhône-Alpes. La ville principale de Rhône-Alpes est Lyon. La distance calculée est donc celle entre Lyon et Londres. Cette procédure a été mise en place pour faire face au nombre important de localisations existantes (ce travail a été réalisé pour l'ensemble des bases PCRD) et le coût en termes de temps qu'aurait demandé la construction d'une matrice de distance entre toutes les localisations possibles.

La prise en compte de l'espace économique

La particularité des programmes cadres est le fait qu'ils couplent à la fois des objectifs en termes d'innovations mais aussi la volonté de l'UE de créer un EER. L'UE impose notamment la présence dans chaque projet d'au moins deux agents de l'UE à 15 membres et/ou d'agents issus d'un nouvel état membre. En effet, les PCRD se déroulent dans un contexte d'adhésion. Malgré ces contraintes, notre échantillon est composé en grande majorité d'agents issus de l'UE à 15 membres et, plus particulièrement, des 5 pays dont les agents sont les plus participatifs que sont l'Allemagne, l'Espagne, la France, l'Italie et le Royaume-Uni. Ces pays sont aussi ceux qui présentent le plus d'agents dans notre échantillon. En outre, nous avons observé dans la partie 1 que les principaux axes de collaborations au sein de l'UE reliaient les capitales économiques de ces pays (cf. section 2.3.4). Il est nécessaire de déterminer quel est l'impact dans les collaborations de l'appartenance d'un agent ou des deux agents à un tel sous-ensemble de pays. Nous identifierons ces pays comme le cœur de l'UE.

Hypothèse 4 : La probabilité que deux agents collaborent est influencée par l'appartenance d'un des agents ou des deux agents à un espace économique

tel que l'Union Européenne ou à un sous-ensemble de pays identifiés comme fortement participatifs tel que le cœur⁵.

La variable $core - UE$ permet de vérifier que l'appartenance des deux agents à un ensemble de pays européens défini comme le cœur de l'Europe favorise ou non les collaborations. Cette variable est dichotomique et est construite de la manière suivante :

$$\begin{cases} core - UE_{ij} = 1 & \text{si } i \text{ et } j \text{ appartiennent tous les deux} \\ & \text{au cœur de l'UE} \\ core - UE_{ij} = 0 & \text{sinon} \end{cases}$$

Les pays du cœur sont les pays qui possèdent le plus d'agents au sein de notre base ainsi que le plus grand nombre de participations. Le graphe L (annexes A.13) représente la place prise par ces agents au sein du réseau de collaboration du 6^{ème} PCRDT. Notre définition du cœur reprend un concept connu en économie géographique puisque le cœur de l'Europe ainsi identifié correspond à la Banane Bleue Européenne sans la Suisse et avec l'Espagne. Le fait de ne pas avoir inclus la Suisse au sein de ce cœur s'explique par le fait qu'elle n'est pas membre de l'UE. Sa présence dans cet ensemble pourrait biaiser les résultats par son statut particulier puisque les agents suisses ne sont pas considérés comme partenaires UE. La présence de l'Espagne se justifie par un rattrapage effectué sur les pays leaders de l'UE ces dernières années. Cette variable permet de tester dans une certaine mesure la proximité géographique (tous les pays étant situés dans une zone géographique restreinte) mais aussi une proximité économique.

De manière similaire, nous avons construit la variable $core1/2$ qui prend la valeur 1 si un des deux agents appartient au cœur de l'UE défini précédemment. Cette variable complète la précédente puisqu'elle permet de tester si l'appartenance d'un des deux agents au cœur de l'UE est un facteur qui joue favorablement sur la probabilité de collaborer ou non. Cette variable est une variable dichotomique définie de la façon suivante :

$$\begin{cases} core1/2_{ij} = 1 & \text{si } i \text{ ou } j \text{ appartiennent au cœur de l'UE} \\ core1/2_{ij} = 0 & \text{sinon} \end{cases}$$

Grâce à ces deux variables, il est possible de déterminer si les agents ont plutôt intérêt à collaborer entre agents du cœur, à privilégier des collaborations entre un agent du cœur et un agent extérieur ou au contraire des

⁵Rappelons que le cœur est composé de l'Allemagne, l'Espagne, la France, l'Italie et le Royaume-Uni.

collaborations entre agents extérieurs au cœur.

Remarques : Aucune variable n'a été créée pour tester l'appartenance des agents à l'UE que ce soit l'UE à 15 ou à 27 membres notamment pour tester les conditions imposées par l'UE dans le financement de ces projets. Pour rappel, certains instruments exigent qu'au sein du projet, il y ait au moins trois agents de nationalité différente et membres de l'UE ou candidats à l'UE ou encore appartenant à des pays associés. Or, notre échantillon est composé de 114 agents qui se répartissent de la manière suivante : 101 agents de l'UE à 15 membres, 6 agents localisés dans des États candidats au moment du 6^{ème} PCRDT et qui ont intégré l'UE depuis et 6 agents localisés dans des pays associés (Suisse et Norvège). De fait, il aurait été peu pertinent de créer une variable qui prenne en compte la contrainte liée aux instruments ou encore qui cherche à déterminer l'impact de l'appartenance des agents à des pays membres de l'UE des 15.

Le résumé de nos hypothèses, les effets attendus de chaque variable ainsi que les statistiques descriptives associées à ces variables sont présentés en annexes A.13 et A.14.

Etant donné les hypothèses que nous souhaitons tester, la matrice des variables explicatives peut se décomposer de la manière suivante dans le modèle :

$$Y_{ik_t}^* = X_{ij_{t-1}}^{R\&D} \beta_{ij}^{R\&D} + X_{ij_{t-1}}^{Net} \beta_{ij}^{Net} + X_{ij_t}^{Geo} \beta_{ij}^{Geo} + \epsilon_{ij_t}$$

les exposants *R&D*, *Net* et *Geo* renvoient respectivement aux variables relatives aux moyens engagés en R&D, aux variables réseaux et aux variables géographiques.

Une différenciation qui peut être introduite dans le cadre des PCRDT tient au fait qu'au sein des programmes cadres des agents de types très différents collaborent. Il peut être alors intéressant de déterminer si les choix de collaboration et les déterminants de ces derniers sont influencés par le type des agents ou si ces déterminants sont homogènes malgré une hétérogénéité des agents. Cette question est l'objet de la section 3.4 du chapitre 3. Dans un premier temps, nous cherchons seulement à vérifier les hypothèses présentées dans cette sous-section et, plus particulièrement, à vérifier que les effets réseaux jouent sur les choix de collaboration.

3.3.2 Premières estimations

Dans cette sous-section, nous présentons les résultats des estimations réalisées pour ce modèle. Afin de vérifier la pertinence de nos estimations et, plus particulièrement, de vérifier que l'hypothèse de nullité de l'ensemble des paramètres est rejetée, nous avons calculé pour chaque modèle le test du score, le test du maximum de vraisemblance et le test de Wald. Ces tests ont à chaque fois conduit à rejeter cette hypothèse de nullité des paramètres. De plus, pour chaque modèle, nous avons calculé deux statistiques rendant compte de la qualité de l'ajustement : l'indice de ratio de vraisemblance (ou pseudo- R^2) de Mac Fadden (1974) et la statistique de Veall et Zimmermann R_{VZ}^2 (1992). Pour l'ensemble des estimations, ces indicateurs sont faibles ce qui est cohérent avec les indicateurs obtenus dans la littérature en économétrie de réseau (Fafchamps, vand der Leij et Goyal, 2010 ; Autant-Bernard et al., 2007).

Le réseau du 5^{ème} PCRDT influence les choix de collaborations réalisés dans le 6^{ème} PCRDT

Pour nos estimations, nous allons procéder par “forward induction” puisque nous commençons par un nombre restreint de variables qui constitue le socle de nos estimations. Le modèle est composé au départ des moyens en R&D dont disposent les différents agents mesurés ici par le coût moyen supporté par les agents dans les projets auxquels ils ont pris part dans le 5^{ème} PCRDT ainsi que de la centralité des agents dans ce même PCRDT mesurée par l'intermédiaire du degré renforcé. Cette mesure présente l'avantage d'être synthétique en termes de centralité. Pour ces deux types de variables (moyenne et différence), nous créons deux variables comme proposé par Fafchamps, van der Leij et Goyal (2010). Ces quatre variables constituent le modèle de base (cf. annexes A.14 modèle 1). Par la suite, les variables nécessaires aux tests des différentes hypothèses seront rajoutées au modèle de base. Les modèles 1 à 3 permettent de tester les hypothèses 1 et 2 relatives respectivement à l'effet réseau et aux moyens en R&D.

Le modèle 1 est le modèle de base. Toutes les variables introduites sont significatives. Nous remarquons que deux agents qui sont centraux dans le 5^{ème} PCRDT sont des partenaires attractifs l'un pour l'autre. En parallèle, nous observons que le signe du coefficient de ΔDR est positif. Ceci signifie que deux effets jouent simultanément : d'un côté plus les agents sont centraux, plus ils ont intérêt à collaborer ensemble mais il existe dans le même temps des incitations pour des agents dissemblables à collaborer. On se retrouve alors dans un schéma de collaboration où les agents centraux collaborent

entre eux mais aussi avec des agents périphériques. Ce type de schéma a été retrouvé dans les études de Goyal, van der Leij et Moraga-Gonzalez (2006). Ces auteurs étudiaient les réseaux de co-auteurs dans le domaine de la recherche en économie. Ils avaient identifiés le schéma suivant : “*the world of economists has been and still is spanned by a collection of inter-linked stars*” (2006, p9). En effet, Goyal, van der Leij et Moraga-Gonzalez (2006) ont montré que les auteurs les plus centraux collaboraient ensemble ainsi qu’avec des auteurs qui comptaient un nombre de co-publications beaucoup plus faible. Il semble donc que ce schéma s’applique aussi aux réseaux de collaborations dans les PCRDT.

Si les coefficients des paramètres des deux variables réseaux sont tous les deux positifs, ce n’est pas le cas des coefficients des paramètres liés aux variables mesurant les moyens en R&D. En effet, le paramètre lié à \overline{COST} est positif alors que celui de $\Delta COST$ est négatif. Cela signifie que la probabilité que deux agents collaborent est d’autant plus importante que les deux agents possèdent des moyens élevés en R&D. Par contre, plus la différence est importante en termes de moyen de R&D, moins la probabilité que ces deux agents collaborent est élevée. Les agents recherchent à collaborer avec des agents qui possèdent des moyens similaires aux leurs. Étant donné que les variables relatives aux moyens en R&D sont calculées sur le 5^{ème} PCRDT, cette variable n’est pas biaisée par le fait que deux agents qui participent à un même projet dans le 6^{ème} PCRDT supportent des coûts identiques.

Les modèles 2 et 3 introduisent chacun une variable supplémentaire pour prendre en compte l’effet réseau dans nos estimations. Ces deux variables, $DG5$ et $PARTCOM$, renvoient à la connaissance que pouvaient avoir deux agents l’un de l’autre dans le 5^{ème} PCRDT. Ces deux variables, si elles prennent en compte toutes les deux un effet réseau, renvoient à des hypothèses légèrement différentes (toutes les deux issues de l’hypothèse 1b).

La variable $DG5$ renvoie à l’hypothèse selon laquelle deux agents étaient proches dans le 5^{ème} PCRDT (en termes de distance au sein du réseau) ont une forte probabilité de collaborer ensemble dans le programme cadre suivant. Cette distance est mesurée par l’inverse de la distance géodésique qui sépare deux agents au sein du réseau ($DG5$). Dans le modèle, on observe que cette variable est significative et a un effet positif confirmant l’hypothèse 1b. Toutefois, il faut rester prudent concernant cette variable. En effet, on observe que le $Pseudo - R^2$ a doublé avec l’introduction de cette variable. Ceci peut s’expliquer par le fait que la variable est pertinente mais aussi par le fait qu’elle introduit un biais. En effet, si $DG5 = 1$ alors les deux agents sont liés dans le 5^{ème} PCRDT. Le tableau F dans l’annexes A.14 relie la variable $LIEN6$ à la variable $DG5$. Le tableau G (cf. annexe A.14) présente la relation

de manière moins détaillée afin d'observer l'impact sur la variable *LIEN6* d'avoir eu un lien dans le 5^{ème} PCRDT (distance géodésique égale à 1), d'avoir un partenaire commun dans le 5^{ème} PCRDT (distance géodésique égale à 2) ou d'être plus distant. Sur le tableau G (annexes A.14), on observe que les binômes qui ont eu un lien dans le 5^{ème} PCRDT collaborent plus fortement par la suite (dans le 6^{ème} PCRDT). En effet, 40,8% des binômes qui ont collaboré dans le 5^{ème} PCRDT renouvellent par la suite cette collaboration contre seulement 15,7% pour les agents qui ont un partenaire commun et pas de collaborations antérieures. Les binômes dont les agents étaient plus distants dans le 5^{ème} PCRDT sont seulement 10,5% à collaborer dans le 6^{ème} PCRDT.

La variable *LIEN5*, qui rend compte du fait que deux agents ont déjà collaboré ensemble, peut entraîner un certain nombre de biais. On peut notamment supposer qu'il existe une hétérogénéité des comportements selon que deux agents ont déjà collaboré ensemble ou non. Ce problème est traité dans la sous-section 3.3.3. Si la distance géodésique est égale à 1, la variable *LIEN5* est aussi égale à 1, il se peut que cette variable introduise aussi un biais dans l'estimation. Il serait donc préférable de retenir une autre variable que *DG5* ou *LIEN5* pour tester l'effet de la proximité au sein des réseaux sur la probabilité que deux agents collaborent. Dans le modèle 3 présenté ci-après, nous utilisons la variable *PARTCOM*.

La variable *PARTCOM* renvoie au fait que si deux agents possèdent des partenaires communs dans le 5^{ème} PCRDT alors la probabilité que ces deux agents collaborent à la période suivante augmente. Cette variable n'est pas liée à la variable *LIEN5* contrairement à la variable *DG*. En effet, le fait d'avoir un partenaire commun n'implique pas automatiquement d'avoir eu un lien dans le 5^{ème} PCRDT. Dans le modèle 3, on observe que cette variable est significative et le paramètre associé est positif. Ceci vient confirmer l'hypothèse 1b et les résultats mis en avant par Gulati (1993, 1998).

Les tableaux H et I (cf. annexe A.14) qui relient la variable *PARTCOM* à la variable *LIEN6* et confirment ce résultat. Sur le tableau H, on met en évidence le fait que posséder au moins un partenaire commun à la période précédente favorise l'apparition d'un lien entre deux agents à la période suivante. En effet, le pourcentage de binômes qui collaborent dans le 6^{ème} PCRDT est deux fois plus élevé pour les binômes qui possédaient au moins un partenaire commun dans le 5^{ème} PCRDT.

Le tableau I présente une décomposition plus fine de cette relation. On observe alors que plus le nombre de partenaires communs dans le 5^{ème} PCRDT augmente, plus le pourcentage de binômes qui collaborent ensuite dans le 6^{ème} PCRDT augmente. Par contre, au delà de 6 partenaires communs, le résultat

est plus contrasté. En effet, au delà de 6 partenaires communs, le pourcentage de binôme qui collaborent par la suite dans le 6^{ème} PCRDT est plus variable. Il est possible qu'il existe un seuil au delà duquel le nombre de partenaires communs influence négativement la probabilité que deux agents collaborent. Il est possible de supposer que si deux agents possèdent un nombre important de partenaires communs, ils peuvent posséder un profil similaire qui ne les incitent pas à collaborer ensemble. Il se peut donc que la relation entre *PARTCOM* et la probabilité que deux agents collaborent dans le 6^{ème} PCRDT suivent une courbe en cloche. Pour tester cette relation, nous estimons un modèle supplémentaire dans lequel nous introduisons le carré de la variable notée *PARTCOM*². Les résultats sont présentés dans le modèle 4 (cf. annexe A.14). Les deux variables sont significatives. Les paramètres associés à *PARTCOM* et *PARTCOM*² sont respectivement positif et négatif. Ils jouent en sens opposé confirmant la présence d'un effet seuil au delà duquel le nombre de partenaires communs n'est plus un facteur incitatif pour collaborer.

Pour les estimations suivantes, nous retiendrons les variables \overline{DR} et ΔDR qui rendent compte de la centralité des agents dans le 5^{ème} PCRDT ainsi que la variable *PARTCOM*. Dans la sous-section suivante, nous introduisons dans le modèle les variables qui nous permettront de tester les effets géographiques.

L'effet de la proximité géographique dans les PCRDT : un effet ambigu

Afin de mesurer le rôle joué par la proximité géographique sur la probabilité pour un binôme de collaborer dans le cadre du 6^{ème} PCRDT, nous pouvons utiliser deux types de variables : des variables continues qui rendent compte de la distance qui sépare les deux agents du binôme, ou des variables dichotomiques qui permettent de situer un agent par rapport à l'autre sans pour autant préciser une distance réelle. L'ensemble de ces variables ainsi que leur combinaison (lorsque c'est possible) sont testés dans les modèles 5 à 13 (cf. annexes A.14).

Les modèles 5 et 6 reposent sur le modèle 3 vu à la sous-section précédente. Le modèle 5 introduit la variable *DISTANCE* qui renseigne sur la distance en kilomètres qui sépare les deux agents du binôme alors que le modèle 6 introduit la variable *TEMPS* qui renseigne le temps nécessaire exprimé en minutes pour effectuer la distance séparant les deux agents par la route. Lorsque l'on observe les estimations obtenues sur les modèles 5 et 6, aucune

des deux variables n'est significative. La proximité géographique ne semble donc pas jouer. On retrouve un résultat similaire de celui de Autant-Bernard et al (2007). Ce résultat peut s'expliquer par les conditions minimales imposées par l'UE dans chacun des projets pour obtenir des financements. Une de ces contraintes oblige chaque projet à compter au moins trois agents issus de pays différents. De plus, il est difficile de déterminer dans quelle mesure le contexte dans lequel se déroule le 6^{ème} PCRDT (adhésion de nouveaux Etats membres) affecte les choix de collaboration. Dans le même temps, la taille moyenne des projets dans le 6^{ème} PCRDT est élevée. De fait, les contraintes peuvent être respectées tout en n'empêchant pas les agents de choisir des partenaires situés à proximité.

Les modèles 7 à 9 introduisent chacun une nouvelle variable par rapport au modèle 3 qui rend compte de la proximité géographique. Les variables introduites sont respectivement *PROXGEO*, *BORDER* et *IDPAYS*.

La variable *PROXGEO* permet de tester si les deux agents du binôme sont localisés dans des pays qui possèdent une frontière commune ou dans le même pays. On observe que dans le modèle 7, cette variable apparaît comme significative. La proximité géographique semble jouer même si cette variable est seulement significative au seuil de 10%. Le paramètre associé est positif ce qui signifie que l'appartenance de deux agents au même pays ou à des pays frontaliers favorise la collaboration. Cette variable montre qu'il existe un effet de proximité géographique particulier au sein des PCRDT même si elle ne rend qu'imparfaitement compte de cette proximité. Le tableau J (cf. Annexes A.14) met en avant le lien entre la variable dépendante et la variable *PROXGEO*. On montre que si deux agents sont localisés dans le même pays ou dans des pays frontaliers, alors la probabilité qu'ils collaborent est de 14,8% contre 12,5%. Ceci explique à la fois le signe positif du paramètre associé à cette variable mais aussi le fait que cette variable ne soit significative qu'au seuil de 10% étant donné la faiblesse de cet écart.

Dans les modèles 8 et 9 (cf. annexes A.14), nous dissociions l'appartenance de deux agents à des pays frontaliers et celle de deux agents au même pays grâce respectivement aux variables *BORDER* et à la variable *IDPAYS*. Les deux variables sont non significatives. Ce résultat peut s'expliquer par le fait que les agents recherchent leurs partenaires dans des pays différents du leur, plus particulièrement dans des pays frontaliers, et qu'en même temps la taille des projets est suffisante pour voir apparaître des collaborations au sein d'un pays et contourner les conditions imposées par l'UE. Les variables *BORDER* et *IDPAYS* qui ne captent qu'un seul de ces deux effets apparaissent alors non significatives.

Au final, la seule variable rendant compte de la proximité géographique qui est significative est *PROXGEO*. Nous conservons cette variable pour la suite des estimations. Nous arrivons à des conclusions proches de l'article de Autant-Bernard et al. (2007) mais différentes de celles de Paier et Scherngell. Ces derniers, étudiant les collaborations dans le cadre du 5^{ème} PCRDT, avaient montré que la distance géographique était un déterminant des choix de collaborations de même que le fait que les agents soient localisés dans le même pays. Ces différences de résultat peuvent s'expliquer de deux manières : premièrement, une manière différente de sélectionner l'échantillon d'agents (cf. introduction de la partie 2) et, deuxièmement, une méthode de calcul différente de la distance (les auteurs retiennent une distance mesurée par l'orthodromie⁶).

L'appartenance à un espace économique influence les collaborations

Un autre déterminant susceptible d'influencer les choix de collaboration est l'appartenance à un sous-ensemble d'agents qui renvoie à un certain espace économique qui présente des propriétés particulières. Dans la partie précédente, nous avons remarqué que les pays qui possédaient les agents les plus participatifs étaient l'Allemagne, l'Espagne, la France, l'Italie et le Royaume-Uni que nous avons définis comme le cœur de l'UE. Il est alors intéressant d'évaluer l'impact sur la probabilité de collaborer de l'appartenance des deux agents du binôme ou de l'un des deux agents au cœur de l'UE. Comme ces agents sont les plus participatifs, on peut supposer que l'appartenance d'au moins un des deux agents aux PCRDT peut influencer les choix de collaboration. De la même manière, l'UE impose qu'au moins deux agents appartiennent à deux pays de l'UE des 15.

Afin de tester cette hypothèse, nous introduisons deux variables au modèle 3 : *Core - UE* et *Core1/2*. Chaque variable est introduite séparément (modèles 10 à 11). Ces variables, même si elles servent à tester l'appartenance à un espace économique, renvoient aussi à un espace géographique. Afin de mesurer les effets de ces deux variables, nous les introduisons dans un premier temps dans le modèle 3 qui ne prend pas en compte les effets de proximité géographique à l'aide de la variable *PROXGEO*. Dans un second temps, nous introduirons simultanément la variable géographique et les variables prenant en compte l'espace économique.

⁶L'orthodromie désigne le chemin le plus court entre deux points d'une sphère, c'est-à-dire l'arc de grand cercle (en géométrie, c'est un cercle tracé à la surface d'une sphère. Il a le même diamètre qu'elle et la divise en deux hémisphères égaux.) qui passe par ces deux points.

L'appartenance des deux agents au cœur de l'UE comme nous l'avons défini précédemment diminue la probabilité de collaborer pour le binôme. Dans le modèle 10, la variable $Core - UE$ est significative (la significativité et le signe des autres variables ne sont pas modifiés). Le paramètre associé à cette variable a un signe négatif ce qui signifie que la localisation des deux agents dans des pays du cœur de l'UE affecte de manière négative la probabilité de collaborer. Le tableau J (cf. annexes A.14) rend compte de la relation entre la variable $Core - UE$ et $LIEN6$. On observe que 10,6% des binômes dont les agents appartiennent tous les deux au cœur de l'UE collaborent dans le 6^{ème} PCRDT contre 15% dans le cas contraire.

Grâce à la seconde variable, il est possible de déterminer si les agents privilégient plutôt une relation entre un agent localisé dans un des pays du cœur et un agent extérieur à ce cœur. Pour vérifier cette assertion, nous avons construit la variable $Core1/2$ qui prend la valeur 1 si un des deux agents est localisé dans le cœur. Les estimations sont données dans le modèle 11. La variable $Core1/2$ apparaît comme non significative. Cela signifie que l'appartenance des deux agents au cœur n'est pas déterminant pour le choix de collaboration. Le tableau K confirme ce résultat. Sur ce tableau, on observe que quand un agent du binôme appartient au cœur, dans 14,3% des cas les agents collaborent alors que ce pourcentage n'est que de 12,5% dans le cas contraire. L'appartenance d'un des agents au cœur n'est pas déterminant pour les choix de collaboration.

Au final, seule la variable $Core - UE$ est significative. Le paramètre associé à cette variable est négatif alors que l'on s'attendait à ce qu'il soit positif montrant que les agents localisés dans le cœur ne collaborent pas entre eux ce qui semble aller à l'encontre des résultats de la partie 1 dans laquelle nous avons montré que les principaux axes de collaborations s'établissaient entre des agents localisés dans les capitales économiques des pays du cœur. Pour autant, cela n'invalides pas les résultats de la partie 1 dans la mesure où l'échantillon que nous étudions est particulier. En effet, il faut rappeler que nous étudions ici seulement les agents des nanotechnologies et, plus précisément, ceux qui ont participé aux deux PCRDT. Afin de vérifier les résultats de la partie 1, il aurait fallu pouvoir conserver tous les agents des domaines Télécommunications et Electronique-Microélectronique.

Pour tester les effets de l'appartenance à un espace économique, nous avons étudié ces variables séparément sans introduire les variables liées à la proximité géographique étudiée au paragraphe précédent. Le modèle 12 (cf. annexes A.14) construit sur la base du modèle 3, introduit simultanément la variable $PROXGEO$, qui prend en compte la proximité géographique, et la variable $Core - UE$, qui prend en compte l'appartenance des agents à un

espace économique.

Dans le modèle 12, toutes les variables ont une influence significative. Les paramètres associés aux variables relatives aux moyens en R&D, à la proximité sociale et à la centralité des agents conservent leur signe et leur effet sur la probabilité de collaborer dans le 6^{ème} PCRDT ainsi que leur degré de significativité. De la même manière, les variables *PROXGEO* et *Core – UE* conservent elles aussi leurs effets.

Les estimations réalisées permettent de déterminer l'influence, positive ou négative, d'une variable sur le choix de deux agents de collaborer ou non dans le 6^{ème} PCRDT. Les valeurs des paramètres ne nous renseignent pas sur l'intensité de cet effet puisque les paramètres obtenus par le biais d'une estimation logit ne sont pas directement interprétables. Pour cela, il nous faut calculer les effets marginaux. C'est l'objet du paragraphe suivant.

Les effets marginaux associés aux différents déterminants des choix de collaboration

Pour calculer ces effets marginaux, nous nous basons sur le modèle 12 qui est le modèle le plus complet. La méthode de calcul de ces effets marginaux est présentée à la section 3.2.2. Les effets marginaux associés à chaque variable sont présentés dans le tableau suivant :

Variables	Type	Complément	Effets marginaux
\overline{DR}	Continue	-	0,0231
ΔDR	Continue	-	0,0236
\overline{COST}	Continue	Exprimée en euros	0,0673
$\Delta COST$	Continue	Exprimée en euros	-0,0550
<i>PARTCOM</i>	Discrète	Données de comptage	0,017
<i>Core – UE</i>	Binaire	-	-0,040
<i>PROXGEO</i>	Binaire	-	0,025

Tableau 38 : Répartition des agents par type.

Les effets marginaux permettent de déterminer l'effet d'une variable, positif ou négatif, sur la probabilité que deux agents collaborent dans le 6^{ème} PCRDT. La probabilité de collaborer est comprises entre 0 et 1.

Pour les variables continues, les effets marginaux semblent peu élevés. Toutefois, il faut prendre en compte dans le calcul des effets marginaux, l'unité utilisée pour exprimer les différentes variables.

Ainsi, les variables \overline{COST} et $\Delta COST$ sont exprimées en euros alors que les moyennes de ces variables sont respectivement égales à 352940 euros et 443206 euros. Le maximum atteint par ces deux variables dépassent les deux millions d'euros. Ces valeurs expliquent la faible valeur des effets marginaux associés à ces deux variables. Une augmentation de 100000 euros des moyens en R&D du binôme augmente la probabilité que les deux agents collaborent de 0,7%. La variable $\Delta COST$ joue dans le sens inverse de la variable \overline{COST} . Une augmentation de la différence entre les deux agents du binôme conduit à une réduction de la probabilité que ces deux agents collaborent. Pour une augmentation de la différence de 100000 euros, la probabilité que deux agents collaborent diminuent de 0,6%.

Les remarques faites pour les variables liées aux moyens en R&D sont valables aussi pour les variables \overline{DR} et ΔDR . Elles sont calculées sur la base du degré renforcé qui est une mesure de centralité qui permet d'évaluer la centralité d'un agent dans un réseau. C'est un indicateur ordinal, la valeur de l'indicateur ne sert qu'à comparer les agents l'un à l'autre. Ainsi, il est difficile d'interpréter les effets marginaux liés à ces variables.

La variable *PARTCOM* permet de compléter la prise en compte des effets réseaux. La probabilité que deux agents collaborent augmentent de 1,7% si le nombre de partenaires communs de deux agents augmente de 1. Toutefois, nous avons montré dans le modèle 4 que la relation entre le nombre de partenaires communs et le fait de collaborer dans le 6^{ème} PCRDT n'est pas linéaire.

Enfin, la localisation de deux agents dans un des pays du cœur diminue la probabilité de collaborer de 4%. La proximité géographique joue en sens opposé mais l'effet marginal associé à cette variable est faible. Si les deux agents sont localisés dans des pays frontaliers ou dans le même pays, la probabilité qu'il collabore augmente de 2,5%. Ceci tendrait à montrer que les agents détournent les contraintes imposées par l'UE et que la proximité géographique l'emporte.

3.3.3 La prise en compte des collaborations antérieures

Dans cette sous-section, nous allons introduire la variable dépendante retardée d'une période dans le modèle *LIEN5* pour prendre en compte les collaborations antérieures, prenant la valeur 1 si deux agents ont collaboré dans le 5^{ème} PCRDT et 0 dans le cas contraire. Nous montrerons les problèmes que peut entraîner l'introduction de cette variable et la méthode utilisée pour estimer l'impact de cette variable en limitant les biais.

Introduction de la variable dépendante retardée et problèmes d'estimation

La variable *LIEN5*, qui rend compte du fait que deux agents ont déjà collaboré ensemble, peut introduire un biais. En effet, l'introduction de notre variable dépendante mais retardée d'une période peut nous laisser supposer qu'il existe un changement dans les choix de comportement une fois qu'une collaboration a déjà eu lieu entre deux agents.

Pour autant, les effets de cette variable sont importants à estimer car ils permettent de contrôler l'importance des collaborations passées pour expliquer les collaborations présentes. En effet, dans la sous-section précédente, nous avons pu observer que les paramètres associés aux variables réseaux étaient significatifs. Les variables *PARTCOM* et *DG* ont un effet positif sur la probabilité de collaborer. Ce résultat permet de valider en partie l'hypothèse 1b dans la mesure où plus les agents sont proches au sein du réseau de la période antérieure, plus la probabilité qu'ils collaborent dans le réseau présent est importante.

Toutefois, pour valider complètement l'hypothèse 1b, il faut évaluer l'impact des collaborations passées sur les collaborations présentes. Cela passe par l'introduction de la variable *LIEN5* dans le modèle. Cette variable peut permettre de tester une dimension nouvelle de la collaboration qui est la relation de confiance qui peut lier deux agents en matière de collaboration en R&D. Étant donné les risques qu'engendrent les activités de R&D, il peut être important pour les agents de s'appuyer sur des partenaires qu'ils connaissent et dont ils connaissent surtout les manières de travailler.

Dans le modèle 13 (cf. annexes A.15), nous introduisons la variable *LIEN5* à la place de la variable *PARTCOM* pour prendre en compte l'effet réseau et tester l'importance des collaborations passées. On observe que cette variable a un effet positif et significatif sur la probabilité de collaborer dans le 6^{ème} PCRDT. Dans le même temps, le *pseudo - R²* évolue peu par rapport au modèle sans la variable *LIEN5*, de même que les valeurs des paramètres du modèle qui restent stables. Le tableau L (cf. annexes A.15) confirme la relation positive qui relie la variable *LIEN5* à la variable dépendante *LIEN6*. Sur ce tableau, nous observons que le pourcentage de binômes qui collaborent dans le 6^{ème} PCRDT est de 41% dans le cas où ces agents ont déjà collaboré dans le 5^{ème} PCRDT alors qu'il n'est que de 12% dans le cas contraire. Une collaboration antérieure est donc un facteur décisif dans les choix de collaboration à la période suivante.

Dans la mesure où le fait d'avoir un lien à la période précédente n'im-

plique pas d'avoir un partenaire commun⁷ et inversement, nous introduisons simultanément les variables *PARTCOM* et *LIEN5* (modèle 14). On observe alors que la variable *PARTCOM* qui avait toujours un effet significatif jusqu'à présent devient non significative. L'ensemble de l'effet proximité sociale est alors capté par la variable *LIEN5*.

Dans notre modèle, il est difficile de tester l'impact que peut avoir l'introduction de la variable dépendante retardée. Nous ne possédons que deux périodes. De fait, il est impossible d'avoir recours à des modèles de panel logit dynamiques qui aurait permis de tester dans le temps l'impact d'une telle variable et la rupture qu'il peut exister en termes de facteurs incitatifs avant et après la première collaboration. Il aurait fallu posséder les données concernant les programmes cadres précédents sachant qu'il n'existe pas une continuité stricte d'un PCRDT à un autre ou alors il aurait fallu partitionner nos deux programmes cadres en plusieurs sous périodes. Cette dernière possibilité aurait entraîné des problèmes en termes de données puisqu'il aurait été difficile de trouver des agents qui participent à toutes les sous-périodes et on se serait trouvé dans le cas d'un panel non cylindré. Pour Raymond et al. (2010) "*When we use the unbalanced panel data set, we need at least three observations over time for some of the firms, two of which need to be consecutive, to be able to identify parameters of the lagged dependent variable*".

Il est donc nécessaire de trouver une solution différente et applicable au modèle logit afin de tester l'importance que peut prendre une collaboration antérieure dans les choix de collaborations présents. Nous utilisons dans le paragraphe suivant un test de changement structurel. Ce dernier permet de tester si le modèle estimé dans cette sous-section s'applique aussi bien au cas où les binômes ont déjà collaboré dans le 5^{ème} PCRDT plutôt que dans le cas contraire.

Des changements structurels liés aux collaborations antérieures

Dans les modèles de régression linéaire, il est supposé que les hypothèses s'appliquent à toutes les observations de l'échantillon. De cette manière, tous les paramètres du modèle sont définis pour l'échantillon total. Toutefois, il est possible de vérifier si certains paramètres varient d'un sous-ensemble de données à l'autre.

⁷Si le nombre d'agents par projet est souvent au minimum de trois agents, 18 des 119 projets financés dans le sous-domaine des nanotechnologies comptent moins de trois agents.

Dans le cas des modèles linéaires, pour tester ces changements structurels, on utilise le test de Chow. Il y a changement structurel si les paramètres ne sont pas égaux pour deux groupes de données issus d'un même échantillon. Le test de Chow sert à vérifier s'il existe une différence dans l'influence d'une variable dépendante entre deux groupes de données, c'est-à-dire si le coefficient est statistiquement différent (Leblond, 2003). Si tel est le cas, les sous-échantillons doivent être estimés séparément. L'estimation réalisée sur l'ensemble de l'échantillon est alors appelée estimation contrainte alors que les estimations effectuées sur chaque sous-échantillon forment l'estimation non contrainte.

Les deux hypothèses que l'on cherche à tester sont les suivantes :

H_0 Les paramètres du modèle sont valables pour toutes les données de l'échantillon.

$$\begin{vmatrix} Y_1 \\ Y_2 \end{vmatrix} = \begin{vmatrix} X_1 \\ X_2 \end{vmatrix} \beta + \begin{vmatrix} \epsilon_1 \\ \epsilon_2 \end{vmatrix}$$

H_A Les paramètres du modèle diffèrent d'un sous-échantillon à l'autre.

$$\begin{vmatrix} Y_1 \\ Y_2 \end{vmatrix} = \begin{vmatrix} X_1 & 0 \\ 0 & X_2 \end{vmatrix} \begin{vmatrix} \beta_1 \\ \beta_2 \end{vmatrix} + \begin{vmatrix} \epsilon_1 \\ \epsilon_2 \end{vmatrix}$$

La matrice Y peut être décomposée en plusieurs sous-échantillons. Pour expliquer nos hypothèses, nous avons pris le cas le plus courant où l'on cherche à montrer qu'il est préférable de découper l'échantillon total en deux sous-échantillons. De manière générale, on découpe le vecteur des variables expliquées en deux matrices colonnes Y_1 et Y_2 auxquelles on associe les matrices contenant les variables explicatives correspondantes X_1 et X_2 et les perturbations ϵ_1 et ϵ_2 . L'hypothèse H_0 décrit le cas où les deux sous-échantillons possèdent les mêmes coefficients pour les variables explicatives. Par contre, l'hypothèse alternative H_A met en avant le fait que chaque sous-échantillon possède des vecteurs de paramètres distincts, ici β_1 et β_2 .

Dans le cas des modèles linéaires, pour tester ces changements structurels, on utilise le test de Chow. Ce test permet de savoir s'il existe une différence significative entre les deux estimations afin de déduire si les sous-échantillons doivent être estimés séparément. Ce test a la forme suivante :

$$Ch = \frac{\epsilon'^* \epsilon^* - \epsilon' \epsilon}{\epsilon' \epsilon} \times \frac{M - 2k}{k} \sim F(k, T - k)$$

avec $\epsilon'^* \epsilon^*$ la somme des carrés des résidus de l'estimation contrainte, $\epsilon' \epsilon$ la somme des carrés des résidus de l'estimation non contrainte (cette somme se décompose en deux, puisqu'on fait la somme de la somme des carrés pour chaque sous-échantillon), M le nombre d'observations et k le nombre de paramètres du modèle sans la constante.

Ce test suit une loi de Fisher. On se fixe un seuil d'erreur (5% en général) qui indique le degré de tolérance pour une erreur. A ce seuil est associée une valeur seuil pour la statistique. Si Ch est supérieur à la valeur seuil alors on rejette H_0 et on en déduit que l'estimation doit être effectuée sur chaque sous-échantillon. A noter que ce test peut être étendu à plus de deux sous-échantillons.

Le test de Chow est valable pour les modèles linéaires mais il peut être adapté aux modèles non linéaires comme le modèle Logit qui est un modèle à choix discrets. Il s'agit alors d'utiliser une contrepartie au test de Chow (Greene, 2005). Ce test consiste à comparer la valeur de la log-vraisemblance du modèle contraint avec la log-vraisemblance du modèle non contraint. Cette dernière est calculée comme la somme de la log-vraisemblance obtenue par l'estimation sur chaque sous-échantillon.

L'adaptation du test de Chow aux modèles de choix binaires passe par l'utilisation du test du ratio de vraisemblance. Nous reprenons la définition de Greene (2005). Le test du ratio de vraisemblance permet notamment de tester les contraintes dans un modèle non linéaire. Dans notre cas, il est utilisé pour tester la stabilité des coefficients du modèle. Nous posons β un vecteur de paramètres à estimer et H_0 une contrainte sur ces paramètres. Soit $\hat{\beta}_{NC}$ le vecteur de paramètres obtenu dans le cas de l'estimation non contrainte et $\hat{\beta}_C$ le vecteur obtenu dans le cas de l'estimation contrainte. Soient \hat{L}_{NC} et \hat{L}_C les fonctions de vraisemblance évaluées à ces deux estimations. Le ratio de vraisemblance est :

$$\lambda = \frac{\hat{L}_C}{\hat{L}_{NC}}$$

Cette statistique est comprise entre 0 et 1 puisque la vraisemblance contrainte est par définition inférieure à la vraisemblance non contrainte. Cette statistique peut se réécrire sous la forme suivante :

$$-2 \ln \lambda = 2(\ln(\hat{L}_{NC}) - \ln(\hat{L}_C))$$

La log-vraisemblance contrainte sera celle du modèle estimé sur l'ensemble de l'échantillon tandis que la log-vraisemblance non contrainte sera la somme des log-vraisemblances obtenues par le même modèle sur chaque sous échantillon (à chaque sous-échantillon correspond une log-vraisemblance). Cette statistique suit une loi du Khi-deux à k degrés de liberté selon le théorème ci-dessous (Greene, 2005) où k est le nombre de contraintes imposées, soit dans notre cas le nombre de paramètres y compris la constante.

Théorème : La distribution limite de la statistique du test de ratio de vraisemblance. Sous la régularité et sous $H0$ (une contrainte sur les paramètres), la distribution en échantillon de grande taille de $-2\ln\lambda$ est une distribution du chi-deux et le nombre de degrés de liberté est égal au nombre de contraintes imposées.

Dans notre cas, l'échantillon est partitionné en deux sous ensembles : les binômes qui ont collaboré dans le 5^{ème} ($LIEN5 = 1$) et ceux qui n'ont pas collaboré dans le 5^{ème} PCRDT ($LIEN5 = 0$). Nous obtenons ainsi deux sous-échantillons qui contiennent respectivement 306 et 6135 observations (cf. annexes A.15 pour les statistiques descriptives de chaque variable pour chaque échantillon). Ensuite, nous effectuons la même régression pour chaque sous-échantillon (modèles 16 et 17 (cf. annexes A.15)) ainsi que sur l'échantillon total de 6441 observations (modèle 15). La régression sur l'échantillon total correspond au modèle contraint alors que celles sur les sous-échantillons correspondent au modèle non contraint. Les résultats sont donnés en annexes A.15.

Dans le cas où les binômes n'ont pas collaboré ensemble dans le 5^{ème} PCRDT, seule la variable *PROXGEO* n'est plus significative. Les autres variables sont toutes significatives et conservent un effet identique à celui observé dans les modèles précédents. Par contre, dans le cas où les agents ont collaboré ensemble dans le 5^{ème} PCRDT, seulement deux variables sont significatives, il s'agit des variables *COUT* et *PROXGEO*. Il faut préciser que pour les modèles 15, 16 et 17 (cf. annexes A.15), l'hypothèse de nullité de l'ensemble des paramètres est rejetée.

La variable *COUT* est la seule significative pour les deux modèles. Quels que soient les liens antérieurs entre les agents, les moyens en R&D constituent un déterminant important des choix de collaboration. La variable *PROXGEO* quand à elle est la seule autre variable qui reste significative pour les agents qui ont déjà collaboré dans le 5^{ème} PCRDT alors qu'elle ne l'est pas dans le cas contraire. L'effet de proximité géographique est donc présent lorsque les agents ont déjà collaboré. Dans le même temps, pour les binômes n'ayant pas collaboré dans le 5^{ème} PCRDT, toutes les variables sont significatives sauf la

variable *PROXGEO*. De plus, on remarque que la variable *CoreUE* a un effet significatif et négatif sur la probabilité de deux agents de collaborer.

On peut alors supposer qu'une proximité géographique entre les agents peut favoriser une relation de confiance entre eux. De plus, la proximité géographique peut permettre de multiplier les contacts en face à face et ainsi d'améliorer l'efficacité économique des agents. L'ensemble de ces paramètres peut favoriser par la suite une reconduction des collaborations. En outre, il est aussi possible que les binômes reconduisent leur partenariat d'un programme à l'autre et modifient les autres partenaires du projet en vue de satisfaire aux conditions imposées par l'UE. Si tel est le cas, les choix de collaborations des binômes qui ont collaboré ensemble dans le 5^{ème} PCRDT doivent être influencés par les variables *DISTANCE* et *IDPAYS* qui rendent compte de la proximité géographique. Pour vérifier cela, nous introduisons dans le modèle 16 la variable *DISTANCE* puis la variable *IDPAYS* à la place de la variable *PROXGEO*. Pour rappel, ces deux variables étaient non significatives lorsqu'elles étaient introduites dans des régressions concernant l'ensemble de l'échantillon. Lorsque l'on introduit ces variables seulement pour les binômes qui ont déjà collaboré dans le 5^{ème} PCRDT, elles sont significatives (au seuil de 1% pour la variable *DISTANCE* et de 10% pour la variable *IDPAYS*). Le paramètre associé à la variable *DISTANCE* est négatif ce qui signifie que plus la distance augmente entre les agents, plus la probabilité qu'ils collaborent diminue. Le paramètre associé à la variable *IDPAYS* est positif, la présence de deux agents dans le même pays favorise donc la collaboration. Ces deux effets étayaient donc l'hypothèse selon laquelle les agents reconduisent d'une année sur l'autre des collaborations avec des partenaires qu'ils connaissent et collaborent avec d'autres agents dans l'objectif de satisfaire aux conditions imposées par l'UE.

Afin de vérifier que ces échantillons doivent être estimés séparément, nous calculons le test de changements structurels appliqués aux modèles logit. Les résultats sont donnés dans le tableau suivant :

	Degré de liberté	Valeur seuil à 5%	Valeur du test
Résultat	7	14,067	151,637

Tableau 39 : Résultat du test de changement structurel 1.

Nous rejetons l'hypothèse nulle selon laquelle l'ensemble de l'échantillon est homogène. Il est plus pertinent de retenir l'hypothèse alternative selon laquelle chaque sous-échantillon doit être estimé séparément. Cela induit donc que la présence de collaborations à la période antérieure modifie les déterminants qui incitent les agents à collaborer.

Tout au long de cette section, nous avons cherché à identifier quels étaient les déterminants qui pouvaient influencer les choix de collaborations des agents. Nous avons retrouvé un certain nombre de résultats présents dans la littérature (Gulati, 1998 ; Fafchamps, van der Leij et Goyal, 2010 ; Autant-Bernard et al, 2007 ; Paier et Scherngell, 2008). Seul l'effet concernant la proximité géographique diffère. De plus, nous avons montré l'importance que prenaient les collaborations passées dans les déterminants des choix de collaborations futures.

Le fait que deux agents aient collaboré à une période antérieure modifie par la suite les comportements des agents, notamment les déterminants qui favorisent la collaboration entre deux agents. De plus, ce résultat valide l'hypothèse 1b selon laquelle la proximité sociale influe sur les choix de collaborations. Pour enrichir nos résultats, il aurait été intéressant de croiser la base de données PCRDT avec d'autres bases qui rendent compte de relations collaboratives du type co-brevets et co-publications afin d'observer si les collaborations dans les PCRDT se retrouvaient dans ces domaines ou si elles étaient antérieures ou postérieures à celles-ci. De plus, le croisement des données PCRDT avec d'autres bases du type brevets et co-publications peut être utile dans un objectif de déterminer quelle est la concrétisation des projets collaboratifs notamment dans un objectif d'évaluation des projets collaboratifs dans les PCRDT. De cette manière, il serait possible de tester si les collaborations au sein des projets se traduisent par des publications et/ou des brevets et quels sont les partenaires du projet qui sont impliqués au final.

Dans la section suivante, nous montrons qu'en dehors des collaborations passées un autre facteur peut entraîner une diversité des déterminants collaboratifs, il s'agit de l'hétérogénéité des agents.

3.4 Une hétérogénéité des comportements au sein des réseaux de collaborations.

Dans cette section, nous montrons tout d'abord que la propension à collaborer est différente selon le type d'agent (section 3.4.1). Cette différence peut s'expliquer par des motivations différentes des agents à collaborer selon leur type. Scherngell et Barber (2009) ont montré que les déterminants des collaborations entre régions pouvaient être variés selon que l'on s'intéressait au réseau des entreprises ou à celui des universités. Leur approche se faisait à un niveau agrégé. Contrairement à ces auteurs, nous cherchons à tester cela au niveau individuel. Deux autres points distinguent l'approche de ces

auteurs de celle retenue dans cette thèse. Premièrement, il est important de chercher quels sont les déterminants qui influent sur les collaborations entre une entreprise d'un côté et un centre de recherche et/ou un agent de l'enseignement supérieur de l'autre. En effet, le développement des collaborations entre agents de type différent est l'un des objectifs principaux des PCRDT. Deuxièmement, il est important de montrer dans quelle mesure la composition du binôme affecte les déterminants des collaborations et si pour chaque type de binôme, il n'est pas plus opportun de réaliser des estimations de manière séparée. Nous testons alors l'effet de l'hétérogénéité des agents sur les choix de collaborations de deux manières : premièrement en introduisant des variables binaires et, deuxièmement, en utilisant un test de changement structurel présenté à la section précédente (section 3.4.2).

3.4.1 Une propension à collaborer différente en fonction du type des agents du binôme

Dans la partie 1, nous avons présenté les différents types d'agents présents dans les PCRDT. Le tableau ci-dessous présente la proportion de ces différents types au sein de notre échantillon.

Type des agents	Nombre	Répartition en %
Centres de recherche	22	19,5
Agents de l'enseignement supérieur	40	35,4
Entreprises	47	40,7
Autres (associations, fédérations,...)	5	4,4
Total	114	100

Tableau 40 : Répartition des agents concernés par les estimations par type.

Les entreprises (PME ou non PME) représentent plus de 40% des agents suivi des agents de l'enseignement supérieur 35,4% et des centres de recherche 19,5%. Le type autres compte seulement 5 agents soit 4,4% de l'échantillon.

Nous avons créé trois variables binaires afin d'observer l'influence que peuvent avoir les différents types d'agents de l'échantillon sur la probabilité de collaborer. Ces trois variables sont : *IND*, *HE* et *RES*. La variable *IND* prend la valeur 1 si les deux agents sont deux entreprises et 0 sinon. La variable *HE* prend la valeur 1 si les deux agents sont des centres de recherche. La variable *RES* prend la valeur 1 si les deux agents sont des agents de l'enseignement supérieur. Le dernier type d'agents est le type Autres qui concernent seulement 5 agents de l'échantillon. Chacune de ces variables est

introduite de manière séparée dans le modèle 12 qui est le modèle qui prend en compte l'ensemble des effets testés dans la section précédente. Les résultats des estimations sont donnés dans l'annexe A.16 (modèles 18 à 20).

La manière dont sont composés les binômes affecte différemment la probabilité que ceux-ci ont de collaborer. Les variables *HE* et *RES* ont un effet significatif et positif sur la probabilité de collaborer (modèle 18 et 19). Si les deux agents du binôme sont deux centres de recherche ou deux agents de l'enseignement supérieur alors la probabilité de collaborer augmente confirmant les résultats de la partie 1 dans laquelle nous avons montré que ces deux types d'agents étaient les plus collaboratifs, même si, pour les besoins de l'estimation, nous avons travaillé ici seulement sur un sous-domaine technologique et non sur les domaines Télécommunications et Electronique-Microélectronique. Dans le même temps, on remarque que la variable *IND* a un effet significatif et négatif. Ceci signifie que, si les deux agents du binôme sont des entreprises, cela influe négativement sur la probabilité de collaborer. Dans le modèle 20 (cf. annexes A.16), nous introduisons simultanément ces trois variables. On observe qu'elles restent toutes les trois significatives et qu'elles conservent leur effet. La composition du binôme influe donc sur la probabilité de collaborer. De plus, tous les agents ne sont pas aussi collaboratifs les uns que les autres.

L'important pour estimer l'effet des différents types d'agents est de créer les catégories d'agents les plus homogènes possibles. Or, comme préciser dans l'annexe A.2, il existe des difficultés à identifier clairement les agents de type centres de recherche et de type agents de l'enseignement supérieur. En effet, une des principales difficultés concernent les universités ou écoles qui possèdent un centre de recherche en leur sein. Dans ces cas là, les déclarations des agents au moment de remplir les dossiers ne sont pas homogènes. Ainsi, des centres de recherche universitaires déclareront soit le centre de recherche (un agent de type centre de recherche est comptabilisé) soit l'université à laquelle le centre de recherche est rattaché (un agent de type agent de l'enseignement supérieur est donc comptabilisé). Au final, il apparaît délicat de conserver ces deux sous-types d'agents.

Afin de prendre en compte ce problème d'identification et de construire des sous-ensembles d'agents homogènes, nous créons un sous-ensemble d'agents qui contient à la fois les centres de recherche et les agents de l'enseignement supérieur. De cette manière, nous créons trois sous-ensembles :

- les entreprises (PME ou non PME)⁸ ;
- les centres de recherche et les agents de l'enseignement supérieur ;

⁸Les centres de recherche des entreprises sont considérés comme des entreprises

– autres.

Ce regroupement permet d'obtenir deux groupes d'agents assez homogènes (le sous-ensemble autres étant à part). En effet, les entreprises ont des impératifs de résultats en termes de profit et une volonté que la recherche débouche rapidement sur un produit commercialisable d'où une recherche plus appliquée. Quant aux agents de l'enseignement supérieur et aux centres de recherche, ils ont surtout des objectifs scientifiques d'où une recherche plus fondamentale.

Type des agents du binôme	Nombre de liens	Nombre de liens potentiels	Pourcentage réalisé
Deux entreprises	101	1035	9,75%
Deux agents de l'enseignement supérieur et/ou centre de recherche	403	1953	20,6%
Mixtes (une entreprise et un centre de recherche et/ou agent de l'enseignement supérieur)	322	2898	11,1%

Tableau 41 : Nombre de liens selon le type des agents du binôme.
La catégorie autres n'a pas été prise en compte dans le tableau.

Dans le tableau ci-dessus, on observe que ce sont les relations entre centres de recherche et/agents de l'enseignement supérieur qui sont les plus nombreuses, aussi bien dans l'absolu qu'en relatif.

De la même manière que nous avons créé les variables *HE* et *RES* pour observer l'effet des types agents de l'enseignement supérieur et centre de recherche sur la probabilité de collaborer, nous créons une variable *HERES* pour tester l'effet du nouveau sous-ensemble d'agents sur cette même probabilité. La variable *HERES* est une variable dichotomique qui prend la valeur 1 si les agents du binôme sont des centres de recherche et/ou des agents de l'enseignement supérieur. Il est important de noter que cette variable renvoie à un plus grand nombre de binômes que les variables *HE* et *RES* dans la mesure où les binômes mixtes de type agents de l'enseignement supérieur et centre de recherche sont pris en compte par la variable *HERES*.

Le modèle 22 de l'annexe A.16 présente les estimations obtenues avec cette variable. On observe que la variable *HERES* a un effet significatif et positif sur la probabilité de collaborer. Dans la mesure où les variables *HE* et *RES* avaient un effet positif, il est normal que la variable *HERES* conserve

cet effet. Le modèle 23 introduit simultanément les variables *HERES* et *IND* dans le modèle. Les deux variables ont un effet significatif. De plus, les deux variables conservent l'effet observé précédemment, à savoir un effet positif pour la variable *HERES* et négatif pour *IND*.

Les résultats suggèrent que les centres de recherche et/ou les agents de l'enseignement supérieur semblent tirer beaucoup plus profit des PCRDT en termes de nombres de collaborations que les entreprises. Ils sont par ailleurs caractérisés par un nombre plus important de liens entre eux. Le résultat peut s'expliquer par le fait suivant : à la différence des agents de l'enseignement supérieur et/ou des centres de recherche, les entreprises peuvent être en concurrence sur le marché. De ce fait, les entreprises peuvent être réticentes à donner des informations aux concurrents lors de la participation à des projets communs.

Il serait possible de tester la validité de cette conjecture en observant que les entreprises qui participent à l'appel d'offre sur les nanotechnologies peuvent être issues de secteurs différents et ne pas être en concurrence. Ainsi, nous retrouvons dans l'échantillon aussi bien des entreprises de l'aéronautique comme EADS que des entreprises de téléphonie comme Nokia, France Telecom, ... Si nous connaissions les codes Nomenclature d'Activités dans la Communauté Européenne (NACE), nous pourrions alors explorer plus en détail la relation entre deux entreprises en regardant notamment si deux entreprises qui possèdent un code NACE identique collaborent ou pas ensemble. Si le fait d'appartenir à un même secteur d'activité joue négativement sur la probabilité de collaborer alors cela viendrait confirmer notre hypothèse.

Une autre raison à la part plus faible de collaborations inter-entreprises pourrait résider dans les contraintes imposées par l'UE aux firmes de rechercher en priorité des partenaires non industriels qui impliqueraient une plus grande selectivité dans les partenariats et un moins grand nombre de partenaires entreprises au sein de chaque projet : *"Particular attention will be given to the strong presence and interaction of innovative enterprises, universities and research organisations in research actions. The integration of education and skills development with research activities will play an important role in increasing European knowledge, in particular in nanosciences and new technologies and opening opportunities for industrial applications"* (Source CORDIS⁹, "Integrating and strengthening the European Research Area" (2004, p11)). Cependant, dans ce cas, ce sont des relations mixtes qui

⁹CORDIS : Service Communautaire d'Information sur la Recherche et le Développement.

devraient être les plus nombreuses. Or, nous nous apercevons que ce n'est pas réellement le cas puisque dans l'absolu ces relations sont aussi nombreuses que les relations entre centres de recherche mais qu'en termes relatifs ce sont les relations entre centres de recherche et/ou agents de l'enseignement supérieur qui occupent une place prépondérante.

Nous avons montré que l'hétérogénéité des agents influe sur la probabilité de collaborer. Pour examiner dans quelle mesure l'hétérogénéité des agents induit une hétérogénéité des comportements qui rendrait des estimations sur des sous-échantillons plus pertinentes qu'une estimation sur l'échantillon total, nous utilisons un test de changement structurel.

3.4.2 Des déterminants différents en fonction du type des agents du binôme

L'objectif est de démontrer par l'outil économétrique un résultat qui peut paraître intuitif : les déterminants qui poussent deux entreprises à collaborer ne sont pas forcément identiques à ceux qui poussent une entreprise et un centre de recherche à collaborer. Dans un second temps, nous déterminerons quels sont les facteurs incitatifs pour chacun des différents binômes.

Nous retenons seulement 109 agents des 114 de l'échantillon utilisé jusqu'à présent. Nous ne retenons pas les 5 agents de type autres. En effet, ce sous-ensemble d'agents est très hétérogène. Il contient aussi bien des cabinets de consultants, des associations, des fédérations, des regroupements d'agents. . . La diversité des agents dans le réseau de collaborations du 6^{ème} PCRDT est représentée au sein du graphe M (annexes A.16).

Ensuite, nous utilisons le test présenté précédemment. Nous partageons l'échantillon en trois sous-échantillons : les binômes constitués uniquement d'entreprises (934 observations), les binômes constitués uniquement de centres de recherche et/ou d'agents de l'enseignement supérieur (1953 observations) et les binômes mixtes (2898 observations).

Les estimations pour chacun des sous-échantillons sont présentées en annexe A.16 (modèles 24 à 27). Le modèle 24 est le modèle contraint. Les modèles 25, 26 et 27 concernent respectivement les binômes composés de centres de recherche et/ou d'agents de l'enseignement supérieur, les binômes mixtes et les binômes constitués d'entreprises. Pour réaliser le test de changement structurel, nous avons besoin de la log-vraisemblance du modèle contraint (modèle 24) et de la log-vraisemblance du modèle non contraint qui est la somme des log-vraisemblances des trois modèles 25, 26 et 27. Le résultat du

test est présenté dans le tableau ci-dessous.

	Degré de liberté	Valeur seuil à 5%	Valeur du test
Résultat	8	15,507	198,43

Tableau 42 : Résultat du test de changement structurel 2.

Le tableau ci-dessus montre qu'il est plus pertinent de retenir l'hypothèse alternative selon laquelle chaque sous-échantillon doit être estimé séparément. Cela induit donc que l'hétérogénéité des agents est source d'hétérogénéité des comportements. Il est important de déterminer maintenant quels sont les facteurs qui influent sur les collaborations de chaque binôme.

Cette hétérogénéité dans les déterminants des choix de collaborations se traduit par des effets différents de chaque variable. La qualité d'ajustement diffère aussi d'un modèle à l'autre. Les mesures de qualité de l'ajustement sont plus élevées pour les paires d'agents du type entreprises. Ensuite, nous remarquons que certaines variables explicatives ne sont pas significatives selon l'échantillon.

Dans un premier temps, nous nous intéressons aux variables réseaux \overline{DR} , ΔDR et $PARTCOM$. Ces variables sont les seules qui sont significatives pour les trois sous-échantillons. Dans les trois sous-échantillons elles conservent un effet positif sur la probabilité de collaborer. Ainsi, quel que soit le type des agents, le fait que deux agents aient un collaborateur commun dans le 5^{ème} PCRDT affecte toujours de manière positive la probabilité que ces deux agents collaborent dans le 6^{ème} PCRDT. De même le schéma collaboratif qui met en avant des collaborations entre agents centraux et entre un agent central et un agent plus périphérique est valable dans les trois modèles. Concernant, les entreprises ce phénomène peut renvoyer à des collaborations entre grandes entreprises d'un côté et des collaborations qui lient PME et non PME. Le tableau ci-dessous présente le rapport entre le nombre de collaborations entre entreprises de chaque type et le nombre potentiel de collaborations entre ces mêmes types d'entreprises.

Type d'entreprises	IND	IND-PME
IND	$15,5\% = \frac{87}{561}$	$2,7\% = \frac{11}{408}$
IND-PME	$2,7\% = \frac{11}{408}$	$9,1\% = \frac{3}{66}$

Tableau 43 : Collaborations entre les différentes formes d'entreprises.

Sur ce tableau, on s'aperçoit que les collaborations se font avant tout entre entreprises non PME. Toutefois, à partir de ce tableau, il est difficile de conclure que le schéma proposé précédemment correspond à la réalité. Il est plus probable qu'il existe une hétérogénéité forte (en termes de tailles, de moyens, . . .) entre les entreprises non PME qui peuvent expliquer les résultats obtenus. Pour pouvoir affiner l'analyse, il faudrait posséder des informations exogènes sur les caractéristiques des entreprises au sens où la source ne serait pas la base PCRDT.

Dans un second temps, nous examinons l'effet des variables relatives aux moyens engagés en R&D. Ces variables apparaissent significatives seulement pour les collaborations où le binôme est mixte. Elles n'influent pas dans le cas des collaborations entre entreprises et entre centres de recherches et/ou agents de l'enseignement supérieur. Lorsque les agents du binôme sont de même type, les choix de collaborations sont influencés par des variables autres que les moyens en R&D comme par exemple une collaboration avec un agent central du réseau. Il serait intéressant de chercher à montrer si les agents de même type qui forment un sous-ensemble homogène recherchent des collaborations avec des agents aux caractéristiques proches des leurs ou au contraire différentes. Il serait également intéressant de tester la proximité sectorielle entre deux agents pour savoir si des agents de même secteur collaborent ou si au contraire les agents recherchent de la diversité au sein des PCRDT. Cette diversité peut être plus ou moins importante avec des agents qui travaillent par exemple dans des secteurs différents mais complémentaires.

Enfin, les variables *PROXGEO* et *Core - UE* ont elles aussi des impacts différents en fonction du sous-échantillon. Elles ne sont pas pertinentes pour expliquer les relations entre firmes. Par contre, la variable *PROXGEO* est significative lorsqu'il s'agit de collaborations mixtes ou de collaborations entre centres de recherche et/ou agents de l'enseignement supérieur. La variable *Core - UE* est significative seulement pour les collaborations entre centres de recherche et/ou agent de l'enseignement supérieur. Cette variable a un effet négatif. Ce résultat peut s'expliquer par le fait que les entreprises, guidées par une logique de profit, sont moins sensibles à la notion de distance à partir du moment où elle savent que la collaboration peut leur apporter les ressources nécessaires, en connaissances et en capital humain, pour mener à bien un projet d'innovation qui peut déboucher sur des résultats concrets et donc sur des profits potentiels. Une hypothèse possible est aussi que des

entreprises distantes l'une de l'autre ne se font pas forcément concurrence par la suite dans la mesure où elles ont un marché qui peut être très localisé ou, au contraire, se rapproche d'entreprises pour intégrer des marchés dans des domaines différents de leurs partenaires potentiels. Une dernière possibilité repose sur un biais dans la déclaration qui veut qu'au lieu d'avoir l'entreprises effectivement impliqué, c'est le siège social qui est déclaré biaisant les résultats sur la proximité géographique. Les centres de recherche étant sensibles à la notion de proximité géographique, lorsqu'ils collaborent avec une entreprise, peuvent chercher à limiter cette distance.

Au sein des PCRDT, il existe des incitations pour favoriser la participation des PME et des agents localisés dans les nouveaux États membres. Par contre, il n'existe pas d'incitations différentes en fonction du type d'agents. Or, nous venons de montrer que les différents types d'agents ne participent pas aux PCRDT pour les mêmes raisons et n'ont pas la même stratégie de collaboration. L'UE devrait donc prendre en compte cette diversité des agents pour améliorer sa politique d'innovation. Par exemple, nous avons vu que la proximité géographique ne jouait pas de manière équivalente pour tous. Les centres de recherche et les agents de l'enseignement supérieur sont plus sensibles à celle-ci. Étant donné cela, l'UE devrait peut être acceptée que des projets de faible taille (trois agents par exemple) ne respectent pas forcément les conditions en termes de nationalité des agents participant aux projets si celui-ci est composé de deux centres de recherche et/ou agents de l'enseignement supérieur et autoriser la participation de deux centres de recherche du même pays dans un projet de faible taille.

Dans ce chapitre, nous avons montré quels étaient les différents déterminants qui influençaient la probabilité que deux agents collaborent dans le 6^{ème} PCRDT.

Nous avons mis en avant que les moyens en R&D (testés par le biais de proxys utilisant les coûts supportés par les agents dans le 5^{ème} PCRDT) influençaient les choix de collaboration. De manière générale, nous avons observé que plus les agents possèdent des moyens importants, plus la probabilité qu'ils collaborent est importante. Dans le même temps, nous avons remarqué que les agents qui possédaient des moyens similaires aux leurs collaboraient entre eux. Le fait de prendre en compte les moyens des agents à la période précédente pour expliquer les collaborations dans le 6^{ème} PCRDT a permis de valider l'hypothèse d'anticipations extrapolatives.

Concernant les variables réseaux, nous avons montré que la place occupée par les agents à la période précédente, mesurée à l'aide de l'indicateur de degré renforcé, jouait un rôle important dans les choix de collaboration.

L'influence positive des deux variables construites à partir de l'indicateur de degré renforcé laisse apparaître un schéma de collaboration où les agents centraux collaborent entre eux tout en nouant des liens avec des agents moins importants. Dans le même temps, nous avons montré que la proximité sociale entre les agents mesurée par le nombre de partenaires communs affectait de manière positive la probabilité de collaborer. De plus, nous avons mis en évidence le rôle essentiel joué par les collaborations réalisées dans le 5^{ème} PCRDT pour expliquer les choix effectués dans le 6^{ème} PCRDT. Le modèle économétrique diffère selon la présence ou non d'une collaboration antérieure entre deux agents.

La seule hypothèse qui n'a pas été entièrement validée est celle relative à la proximité géographique. Nous avons montré que la distance séparant deux agents n'était pas une variable significative sauf pour les agents ayant déjà collaboré à une période antérieure. Toutefois, les choix de collaborations dans le 6^{ème} PCRDT sont influencés par le fait que deux agents soient localisés dans un même pays ou dans des pays frontaliers. La proximité géographique semble avoir un effet limité qui peut s'expliquer par les différentes conditions imposées par l'UE dans les projets ainsi que par l'incitation à coopérer avec des agents localisés dans les nouveaux États membres. Étant donné la taille des projets, il est possible pour un agent de respecter ces conditions tout en collaborant avec des agents qui lui sont proches.

Enfin, l'hétérogénéité des agents influence elle aussi les déterminants des collaborations. Nous avons montré que lorsqu'on s'intéresse aux collaborations entre firmes, seules les variables réseaux jouent un rôle dans les choix de collaborations. Les variables liées aux moyens en R&D ne sont pas un déterminant significatif pour expliquer les collaborations entre firmes mais le deviennent lorsqu'on explique les choix de collaboration entre une firme et un centre de recherche par exemple. Lorsque les agents cherchent à collaborer avec des agents de type identique, les moyens financiers n'entrent pas en jeu. Enfin, la proximité géographique est un déterminant important pour les choix de collaboration des centres de recherche et des agents de l'enseignement supérieur. Quelle que soit la manière de tester l'influence de l'hétérogénéité des agents (variables binaires et test de changement structurel), nous arrivons à la conclusion que le type des agents influe sur les choix de collaboration.

Il est possible d'envisager plusieurs possibilités pour améliorer les modèles présentés dans ce chapitre soit à partir des données sur les PCRDT seulement ou en complétant nos informations par des données extérieures aux PCRDT.

La première possibilité serait de disposer de plus de données sur l'ensemble des PCRDT, notamment d'un plus grand nombre de programmes et de prendre en compte un plus grand nombre de thématiques. De cette ma-

nière, il serait possible d'utiliser des modèles de panel dynamique et d'inclure des variables relatives aux thématiques et domaines technologiques afin de prendre en considération l'activité des agents en fonction des domaines auxquels ils participent. L'utilisation des programmes cadres précédents et d'un plus grand nombre de thématiques demanderait alors un travail d'harmonisation entre les PCRDT.

La deuxième possibilité serait de compléter les informations que nous possédons sur les agents en croisant la base PCRDT avec d'autres bases de données afin d'obtenir des informations supplémentaires sur les caractéristiques de ces agents (moyens en R&D, nombre de chercheurs, chiffre d'affaires, secteurs d'activité,...) et sur les autres types de collaborations tels que les co-brevets et/ou les co-publications. Pour enrichir nos résultats, il aurait été intéressant d'utiliser d'autres bases qui rendent compte de relations collaboratives du type co-brevets et co-publications afin d'observer si les collaborations dans les PCRDT se retrouvaient dans ces domaines ou si elles étaient antérieures ou postérieures à celles-ci.

La dernière possibilité évoquée ouvre de nouvelles pistes de travail dans la mesure où les collaborations dans les PCRDT ne seraient plus l'objet d'étude en lui-même mais un des éléments susceptibles d'expliquer d'autres formes de collaborations comme les co-publications et les dépôts de brevets. Ces collaborations peuvent être vues comme des suites logiques aux collaborations au sein des PCRDT. Pour cela, le croisement des données PCRDT avec d'autres bases du type brevets et co-publications peuvent être utiles pour déterminer quelle est la concrétisation des projets collaboratifs en matière d'innovation. La difficulté serait alors de trouver des bases européennes permettant d'obtenir des informations homogènes d'un pays à un autre et demanderait d'améliorer le processus d'identification des agents. De plus, ce travail permettrait de compléter l'analyse des PCRDT en s'intéressant notamment à l'impact des collaborations au sein des PCRDT sur l'innovation au travers de deux mesures que sont les brevets et les co-publications.

Pour compléter notre étude des collaborations au sein des PCRDT, nous avons évoqué des améliorations potentiels des modèles du chapitre 3. Il est aussi possible de compléter l'analyse en utilisant d'autres cadres complémentaires à celui de l'étude de l'existence d'une collaboration entre deux agents. En vue de cet objectif, deux autres pistes de travail sont explorées et constituent l'objet des deux chapitres suivants. Le chapitre 4 s'intéresse non pas à la formation des collaborations dans le 6^{ème} PCRDT mais à l'intensité de ces collaborations. En effet, jusqu'à présent, nous avons étudié la formation ou non d'une collaboration entre deux agents. Toutefois, nous avons remarqué dans la partie 1 que certains agents collaboraient de manière plus intense. Il

est alors intéressant de déterminer si les facteurs incitatifs identifiés dans le chapitre 3 permettent aussi d'expliquer l'intensité des collaborations. Enfin, dans le chapitre 5, nous nous intéressons non pas à la formation de liens bilatéraux mais à la formation de triades.

Chapitre 4

Expliquer l'intensité des liens

Jusqu'à présent, nous nous sommes intéressés à la détermination des facteurs qui expliquent que deux agents collaborent ou non dans le cadre des PCRDT. Nous avons ainsi montré que les moyens en R&D des agents, la centralité des agents, la proximité géographique et sociale étaient des déterminants de ces choix. Nous nous sommes intéressés à l'existence ou non d'une collaboration entre deux agents.

Dans la partie I, nous avons montré au travers des réseaux valués que certains agents collaboraient plus d'une fois l'un avec l'autre. Il est alors intéressant de se demander si les déterminants identifiés dans le chapitre 3 permettent aussi d'expliquer l'intensité de ces collaborations. Si c'est le cas, il faut alors observer si certains déterminants ne deviennent pas plus importants lorsque l'on s'intéresse à l'intensité des collaborations et non plus seulement à l'existence de ces dernières.

L'étude de l'intensité des liens requiert des modèles d'estimation différents de ceux présentés aux chapitres précédents. En effet, le modèle logit n'est pas valable dans le cas de données de dénombrement. La section 4.1 présente le modèle de Poisson qui est le modèle couramment utilisé pour étudier ce type de données. Dans cette section, nous chercherons à déterminer si les hypothèses présentées dans le chapitre précédent sont valables lorsque l'on s'intéresse à l'intensité des collaborations ? Ensuite, dans la section 4.2, nous montrerons les limites du modèle de Poisson et les corrections qu'il est nécessaire d'apporter au modèle pour améliorer les estimations. Enfin, dans la section 4.3, nous présenterons le modèle binomial négatif qui relâche certaines hypothèses du modèle de Poisson afin de permettre une meilleure estimation.

4.1 Présentation du modèle de Poisson

Tout d'abord, nous présentons la variable expliquée du modèle qui n'est plus l'existence ou non d'une collaboration entre deux agents mais l'intensité de la collaboration (sous-section 4.1.1). Puis, nous présenterons le modèle de Poisson qui est le modèle couramment utilisé pour étudier les données de dénombrement (sous-section 4.1.2). Enfin, nous réaliserons les premières estimations afin de déterminer les facteurs explicatifs de l'intensité des collaborations entre agents (sous-section 4.1.3).

4.1.1 Distribution de l'intensité des collaborations

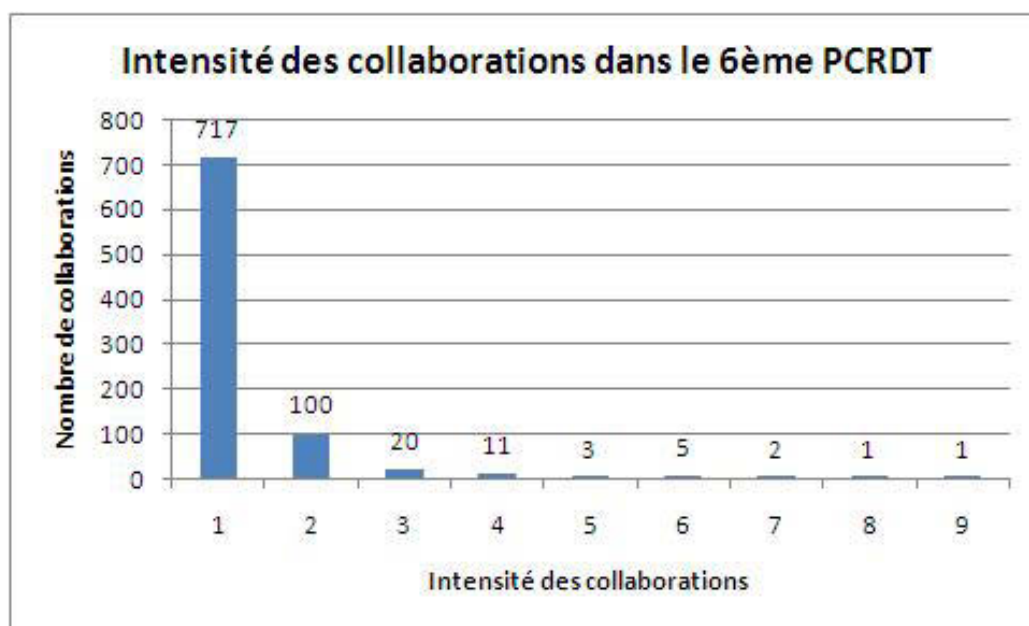
Dans le chapitre précédent, nous avons étudié l'existence ou non d'une collaboration entre deux agents. La variable expliquée était une variable dichotomique qui prenait la valeur 1 si deux agents collaboraient au moins une fois dans le 6^{ème} PCRDT. Dans ce chapitre, nous cherchons à expliquer l'intensité de ces collaborations. L'ensemble des agents retenus est identique à celui du chapitre précédent. L'échantillon est donc composé de 114 agents et il est donc possible de former 6441 binômes potentiels.

La variable expliquée de notre modèle est donc égale à :

$y_{ij} = k$ si les agents i et j collaborent k fois ensemble dans le 6^{ème} PCRDT

La variable expliquée n'est donc plus une variable dichotomique mais une variable de dénombrement qui peut prendre les valeurs $k = 0, 1, 2, 3, \dots$ tel que $k \in N$.

Dans la mesure où l'échantillon est identique, sur les 6441 collaborations potentielles, seules 580 ont une valeur supérieure à 0 et se répartissent de la manière suivante :



Graphique 33 : Intensité des collaborations supérieur ou égal à 1 dans le 6^{ème} PCRDT.

Sur le graphique ci-dessus ne figure pas le nombre de collaborations potentielles non réalisées (intensité égale à 0) pour des soucis de lisibilité. Le nombre de collaborations d'intensité 0 est égal à 5581. La distribution de l'intensité des liens se caractérise donc par une majorité de 0 et de faibles valeurs. Le nombre de binômes qui collaborent de manière étroite dans de nombreux projets est très faible. Le tableau ci-dessous présente l'intensité moyenne des collaborations dans le cadre du 6^{ème} PCRDT.

Indicateurs calculés sur l'échantillon	Intensité des liens 6 ^{ème} PCRDT	Intensité des liens 5 ^{ème} PCRDT
Moyenne	0,171	0,054
Ecart-type	0,536	0,264
Coefficient de variation ¹	3,134	4, 879

Tableau 44 : Intensité des collaborations dans le domaines des nanotechnologies dans les deux programmes cadres.

Sur ce tableau, on observe que l'intensité des collaborations pour notre échantillon est faible dans les deux programmes cadres (mais plus élevé que si l'on considérait le réseau de l'ensemble des collaborations). Ceci est dû à la présence importante de 0 dans les observations, c'est-à-dire d'un nombre important de binômes qui ne collaborent pas. L'écart-type est élevé relativement à

la moyenne. Il existe donc une dispersion importante des observations dans l'échantillon. L'intensité moyenne des collaborations a augmenté entre le 5^{ème} PCRDT et le 6^{ème} PCRDT. Par contre, on remarque que la dispersion était plus importante dans le 5^{ème} PCRDT. Les graphes N et O (annexes A.17) représentent l'intensité des collaborations dans les deux PCRDT.

Le nombre important d'observations qui prennent la valeur 0 associé au fait que la variable est discrète font que les modèles linéaires habituels et les moindres carrés ne sont pas optimaux dans ce cadre là. C'est pourquoi nous utiliserons le modèle de Poisson pour étudier l'intensité des collaborations.

4.1.2 Présentation du modèle de Poisson

La présentation du modèle de Poisson retenue dans cette section est celle de Greene (2005).

Le modèle de régression de Poisson est le modèle couramment utilisé lorsque la variable est une variable de dénombrement, c'est-à-dire une variable qui peut prendre pour valeur des entiers naturels. De plus, la régression de Poisson permet aussi de prendre en compte la prépondérance des zéros et des faibles valeurs de la variable dépendante.

L'équation principale du modèle de Poisson de paramètre λ_{ij} est la suivante :

$$Prob(Y = y_{ij}|x_{ij}) = \frac{e^{-\lambda_{ij}} \lambda_{ij}^{y_{ij}}}{y_{ij}!}$$

avec ij qui renvoie à un des 6441 binômes, y_{ij} la variable dépendante, x_{ij} les variables explicatives et le paramètre λ_{ij} qui est le paramètre de la loi de Poisson. La formulation la plus courante pour exprimer λ_{ij} est le modèle log-linéaire suivant :

$$\ln(\lambda_{ij}) = x'_{ij}\beta$$

La particularité du modèle de Poisson réside alors dans le fait que :

$$E[y_{ij}|x_{ij}] = Var[y_{ij}|x_{ij}] = \lambda_{ij} = e^{x'_{ij}\beta}$$

Afin d'estimer la valeur des paramètres, il faut recourir au maximum de vraisemblance ou plus exactement chercher à maximiser la log-vraisemblance ($\ln L$) qui est :

$$\ln L = \sum_{i=1}^n [-\lambda_{ij} + y_{ij}x'_{ij}\beta - \ln y_{ij}!]$$

Les équations de vraisemblance sont :

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^n (y_{ij} - \lambda_{ij})x_{ij} = 0$$

Le hessien est :

$$\frac{\partial^2 \ln L}{\partial \beta \beta'} = - \sum_{i=1}^n \lambda_{ij} x_{ij} x'_{ij}$$

Le hessien est défini négatif pour tout x_{ij} et β ce qui nous assure d'avoir un maximum. La validité du modèle peut être testée à l'aide des trois tests couramment utilisés que sont le ratio de vraisemblance, le multiplicateur de Lagrange et le test de Wald.

Les paramètres estimés (β) correspondent à l'effet des variables explicatives sur le logarithme de la variable dépendante. En effet, nous avons vu que :

$$\ln(\lambda_{ij}) = x'_{ij}\beta$$

On peut réécrire cette expression de la manière suivante :

$$\lambda_{ij} = e^{(x'_{ij}\beta)}$$

Afin d'obtenir l'effet provoqué par l'augmentation d'une unité de la variable explicative (toutes choses étant égales par ailleurs), il faut donc prendre l'exponentiel du paramètre associé à la variable explicative qui nous intéresse.

Les hypothèses que nous cherchons à tester sont les mêmes qu'au chapitre précédent sauf qu'au lieu de tester l'impact des différentes variables sur la probabilité de collaborer, nous testons l'impact de ces variables sur l'intensité des collaborations. Les variables utilisées pour l'estimation sont donc identiques à celles du chapitre précédent. Les résultats des estimations obtenues à partir du modèle de Poisson sont présentés dans la sous-section suivante.

4.1.3 Des déterminants similaires pour expliquer l'existence et l'intensité des collaborations

Nous cherchons à montrer si les facteurs qui déterminent les choix de collaborations ont aussi un impact sur leur intensité. Pour cela, nous avons donc retenu le même échantillon qu'au chapitre précédent et les variables utilisées sont identiques. Les résultats sont présentés dans l'annexe A.17 (modèles 28 et 29).

La seule distinction entre le modèle 28 et le modèle 29 est la variable qui mesure la proximité sociale. Dans le modèle 28, nous avons utilisé la variable *PARTCOM* alors que dans le modèle 29 nous utilisons la variable *LIEN5*. Cette dernière n'est plus la variable dépendante retardée puisque la variable explicative est l'intensité de la collaboration et non son existence.

La première remarque que l'on peut faire est que l'ensemble des variables ont un effet significatif et les paramètres associés conservent leur effet positif ou négatif. Le fait d'étudier non plus l'existence d'une collaboration entre deux agents mais l'intensité de cette dernière n'influe pas sur l'effet des variables explicatives. L'effet de chaque variable explicative sur la valeur de la variable expliquée est donné dans le tableau suivant :

Variable	Modèle avec <i>PARTCOM</i>	Modèle avec <i>LIEN5</i>
\overline{DR}	1,002	1,002
ΔDR	1,002	1,002
\overline{Cost}	2,550	2,550
$\Delta Cost$	0,568	0,556
<i>Partcom</i>	1,159	
<i>Lien5</i>		3,864
<i>coreUE</i>	0,664	0,677
<i>Proxgeo</i>	1,349	1,316

Tableau 45 : Effet des variables explicatives retenues dans le modèle de Poisson.

Sur le tableau ci-dessus, on observe que l'effet des différentes variables diffère peu selon que l'on utilise la variable *PARTCOM* ou la variable *Lien5*. Par contre, on observe une forte différence de l'impact de ces deux variables sur l'intensité des collaborations alors qu'elles rendent toutes les deux compte de la proximité sociale. Le fait que deux agents aient eu un lien dans le 5^{ème} PCRDT implique que le nombre de collaborations entre ces deux agents est presque 4 fois supérieur par rapport au cas où ils ne collaborent pas dans le 5^{ème} PCRDT. Pour compléter l'estimation, nous avons calculé l'intensité moyenne des collaborations étant donné la présence ou non d'un lien dans

le 5^{ème} PCRDT. Les résultats sont donnés dans le tableau M (cf. annexes A.17). Dans ce tableau, nous remarquons que le nombre de collaborations entre deux agents dans le 6^{ème} PCRDT est 4,5 fois plus élevé en moyenne si ces agents ont collaboré ensemble dans le 5^{ème} PCRDT. Les résultats de l'estimation confirment donc cette approche statistique. L'effet de la variable *PARTCOM* joue dans le même sens. L'augmentation d'une unité du nombre de partenaires communs multiplie par 1,16 le nombre de collaborations dans le 6^{ème} PCRDT. Les deux autres variables relatives au réseau \overline{DR} et ΔDR sont significatives mais ont un effet plus faible sur la valeur prédite du nombre de collaborations.

Lorsque l'on se réfère aux variables relatives aux moyens en R&D, les effets sont importants (modèles 28 et 29). La valeur maximale prise par ces deux variables est respectivement de 2,4 millions d'euros et 2 millions d'euros ce qui explique que l'augmentation d'une unité, soit un millions d'euros, ait de fortes implications sur le nombre de collaborations. Le tableau N (cf. annexes A.17) montre la valeur moyenne des deux variables en fonction de l'intensité de la collaboration dans le 6^{ème} PCRDT. On observe que les variables liées aux moyens en R&D augmentent avec l'intensité des collaborations même si cette relation n'est pas strictement croissante. Il semble donc que plus les agents possèdent des moyens importants plus ils collaborent de manière intense dans le 6^{ème} PCRDT. Si l'on se réfère au tableau ci-dessus, la relation entre la différence en termes de moyens en R&D et l'intensité des collaborations n'est pas aussi simple que le laisse penser nos estimations. En effet, la relation n'est pas strictement décroissante. Il est donc plus délicat de conclure sur l'effet de cette variable.

Enfin, le fait que les agents soient localisés dans le même pays ou un pays voisin a un impact important sur le nombre de collaborations prédites. En effet, ce dernier est 1,3 fois plus important si les deux agents sont situés proches géographiquement. Le tableau O (cf. annexes A.17) présente l'intensité moyenne des collaborations selon la valeur de la variable *Proxgeo*. La proximité géographique joue de manière positive sur l'intensité moyenne des collaborations. Dans le même temps, l'appartenance des deux agents du binôme au cœur de l'UE influe toujours négativement sur l'intensité des collaborations.

Remarque : Les variables qui s'étaient révélées non significatives dans le chapitre précédent comme la variable *Distance* ont été testées. Pour l'ensemble des variables construites dans le cadre de la thèse, nous avons réalisé un test de Wald de type III. Ce test sert à mesurer l'apport d'une variable

pour un modèle en comparant ce modèle avec lui même mais sans la variable testée. Si le test est non significatif alors cela signifie que l'apport marginal de la variable au modèle est nul. Pour toutes les variables qui figurent dans les modèles 28 et 29 le test est significatif. Pour les variables *Distance* et *Temps*, les tests ont été non significatifs. De fait l'effet proximité géographique semble limité.

Le principal problème que l'on rencontre lorsque l'on utilise un modèle de Poisson est la possibilité d'être confronté à de la surdispersion. La surdispersion a pour effet de surestimer les statistiques des tests et donc conduit à considérer des variables significatives alors qu'elles ne le sont pas. Dans la section suivante, nous définissons plus précisément la surdispersion et nous présentons les moyens de repérer ce phénomène ainsi que de le corriger.

4.2 Limites du modèle de Poisson et corrections

La surdispersion est un phénomène qui apparaît notamment dans les estimations à partir des modèles de Poisson sur des données de comptage. Il est donc nécessaire de tester la présence ou non de surdispersion dans le modèle. La surdispersion correspond au fait qu'il existe dans le modèle une plus grande variabilité que celle supposée au travers de la distribution choisie. Cette variabilité implique que la restriction "espérance de la variable expliquée égale variance" apparaît comme trop forte. Généralement, la variance est supérieure à l'espérance d'où des problèmes de surdispersion. Celle-ci peut être causée par l'omission de variables explicatives, par l'existence d'erreurs dans les variables explicatives ou par une mauvaise spécification du modèle.

Le modèle de Poisson a été critiqué du fait de l'hypothèse forte que l'espérance et la variance de la variable expliquée sont considérées comme identiques. La surdispersion conduit à une sous-estimation des écarts-types qui impliquent une surestimation de la valeur des tests qui entraîne, à son tour, une augmentation de la significativité. De fait, certaines variables apparaissant comme significatives dans nos régressions peuvent en fait ne pas l'être. Il est donc important de repérer ce phénomène (sous-section 4.2.1) et, s'il est présent, de le corriger (sous-section 4.2.2).

4.2.1 Test de surdispersion

Dans cette sous-section, nous présentons deux moyens de repérer la surdispersion dans le modèle de Poisson : premièrement, à l'aide du critère de

Pearson et, deuxièmement, à l'aide d'une procédure de Cameron et Trivedi (1990).

Pearson Chi-square

Le premier critère est un indicateur appelé le Pearson Chi-Square (*PCS*) qui évalue la qualité de l'ajustement. Il détermine si le modèle est surdispersé ou au contraire sous-dispersé. Le critère de Pearson se calcule de la manière suivante :

$$\chi^2 = \sum_{ij=1}^n \left(\frac{(y_{ij} - \lambda_{ij})^2}{\lambda_{ij}} \right)$$

avec y_{ij} la valeur observée (l'intensité de la collaboration entre les agents i et j), λ_{ij} la valeur prédite par le modèle et n le nombre de binômes (ici 6441). A partir de χ^2 , on obtient *PCS* de la manière suivante :

$$PCS = \frac{\chi^2}{dl}$$

avec dl le nombre de degré de liberté du modèle ($dl = n - p$ avec p le nombre de paramètres plus la constante).

Ce critère s'interprète de la manière suivante. Si cet indicateur est supérieur à 1 alors on se retrouve dans le cas où le modèle est surdispersé. S'il est inférieur à 1 alors on est dans le cas où le modèle est sous-dispersé.

Dans le cas des modèles présentés dans l'annexes A.17, nous obtenons les résultats suivants :

Modèles	χ^2	dl	<i>PCS</i>
Modèle avec <i>Partcom</i>	15085,14	6433	2,345
Modèle avec <i>Lien5</i>	9973,62	6433	1,55

Tableau 46 : Indicateurs de surdispersion.

Dans les deux cas, nous observons que le modèle est surdispersé. Ce résultat confirme celui obtenu grâce aux statistiques descriptives à la section 4.1.1. Il est donc important de corriger cette surdispersion pour évaluer au plus juste la significativité des paramètres. Toutefois, avant de corriger cette surdispersion, nous testons la présence de surdispersion par un autre critère, celui de Cameron et Trivedi (1990) pour confirmer le résultat acquis. C'est l'objet du paragraphe suivant.

Critère de Cameron et Trivedi

Le second critère repose sur une procédure différente. Cameron et Trivedi (1990) cherchent à tester une hypothèse alternative à celle du modèle de Poisson qui suppose que l'espérance de la variable expliquée est égale à sa variance. Le critère de Cameron et Trivedi permet de comparer les deux hypothèses suivantes :

$$\begin{aligned} H_0 : \quad & \text{var}(y_{ij}) = E(y_{ij}) \\ H_A : \quad & \text{var}(y_{ij}) = E(y_{ij}) + \alpha(E(y_{ij})) \end{aligned}$$

Le test de Cameron et Trivedi consiste alors à déterminer laquelle de ces deux hypothèses il faut retenir. Pour cela, Cameron et Trivedi proposent la procédure suivante. Dans un premier temps, il faut construire la variable z_{ij} qui est définie de la manière suivante :

$$z_{ij} = \frac{(y_{ij} - \lambda_{ij})^2 - y_{ij}}{\lambda_{ij}\sqrt{2}}$$

Puis, dans un second temps, il faut régresser cette variable soit sur un terme constant soit sur λ_{ij} sans terme constant. Ensuite, il est nécessaire d'utiliser un test de Student afin de déterminer si la variable ainsi créée est significative. Si la variable est significative alors il faut conserver l'hypothèse nulle, dans le cas contraire, il faut la rejeter. Dans ce cas, la surdispersion est présente dans le modèle.

Le critère est appliqué aux modèles 28 et 29 de l'annexe. Dans les deux cas, les tests conduisent à rejeter l'hypothèse nulle et donc à conserver l'hypothèse alternative. Dans ces deux modèles, on note la présence de surdispersion qui peut influencer sur les tests de significativité. Il existe donc un risque qu'une ou plusieurs variables de ces modèles qui sont toutes significatives puissent en réalité être non significatives. Il est donc important de corriger l'effet de la surdispersion afin d'avoir des estimations plus robustes. Une première méthode pour corriger la surdispersion est présentée dans la sous-section suivante.

4.2.2 Corrections de la surdispersion

Une méthode permettant de corriger la surdispersion est d'utiliser non plus un modèle de Poisson mais un modèle de quasi Poisson. Le modèle n'utilise plus une fonction de vraisemblance mais une fonction de quasi vraisemblance. L'utilisation d'un modèle de Poisson modifié, c'est-à-dire qui prend en compte le fait que la variance et l'espérance peuvent être différentes, per-

met d'intégrer la dispersion du modèle. Pour cela, on introduit un paramètre, appelé paramètre d'échelle ou paramètre de surdispersion. On a alors :

$$\begin{cases} E(y_{ij}) = \lambda_{ij} \\ V(y_{ij}) = \tau \lambda_{ij} \text{ avec } \tau \geq 1 \end{cases}$$

Le paramètre τ est estimé par $\frac{\chi^2}{dl}$. Cela correspond au critère de Pearson (*PCS*) présenté dans la sous-section précédente.

Dans les modèles 28 et 29 en annexes A.17, le paramètre d'échelle est fixé à 1 par défaut. En effet, dans le modèle de Poisson, on suppose que l'espérance de la variable expliquée est égale à la variance. Dans le cas du modèle de Poisson modifié, le critère de Pearson est utilisé comme estimation du paramètre afin de tenir compte de la surdispersion. Le paramètre est égal à la racine carrée du critère de Pearson.

L'estimation des paramètres n'est pas affectée par cette modification du modèle. Cette modification n'affecte pas non plus les estimations de y_{ij} . Par contre, la matrice de covariance est multipliée par ce paramètre. Les écarts-types du modèle sont augmentés de $\tau^{\frac{1}{2}}$ (Agresti, 1990). De fait, si les estimations du modèle ne sont pas modifiées, leur significativité est affectée par ce changement. L'écart-type étant plus élevé, la valeur du test diminue ce qui peut conduire à rendre certaines variables non-significatives.

Lorsque l'on utilise une fonction de quasi vraisemblance, au lieu de spécifier une distribution de probabilité pour les données, on spécifie une fonction qui relie la variance à l'espérance et qui dans le même temps fait intervenir le paramètre d'échelle.

La fonction de quasi vraisemblance utilisée pour estimer les paramètres est la suivante :

$$\mathcal{Q}(\lambda, y) = \sum_{ij} Q_{ij}(\lambda_{ij}, y_{ij})$$

On a $Q_{ij}(\lambda_{ij}, y_{ij}) = \int_{\lambda_{ij}} \frac{y_{ij} - \lambda_{ij}}{V(\lambda_{ij})} d\lambda_{ij} + f(y_{ij})$

avec $V(\lambda_{ij})$ la fonction de variance et $f(y_{ij})$ une fonction de y_{ij} indépendante de λ_{ij} (McCullagh et Nelder, 1989). La fonction de variance est donnée par $V(\lambda_{ij}) = Var(y_{ij}) = \tau \lambda_{ij}$. Si l'on développe cette dernière équation, on obtient le résultat suivant :

$$\begin{aligned} Q_{ij}(\lambda_{ij}, y_{ij}) &= \int_{\lambda_{ij}} \frac{y_{ij} - \lambda_{ij}}{\tau \lambda_{ij}} d\lambda_{ij} + f(y_{ij}) \\ &= \frac{1}{\tau} \int_{\lambda_{ij}} \frac{y_{ij} - \lambda_{ij}}{\lambda_{ij}} d\lambda_{ij} + f(y_{ij}) \\ &= \frac{1}{\tau} (y_{ij} \ln(\lambda_{ij}) - \lambda_{ij} + C) \end{aligned}$$

A partir de cette équation, on remarque que la maximisation de cette fonction est équivalente à celle de la vraisemblance du modèle de Poisson dans la mesure où $\frac{1}{\tau}$ et C sont indépendants des λ_{ij} .

Les modèles 30 et 31 (cf. annexes A.17) sont des répliques des modèles 28 et 29 à la différence que nous avons utilisé un paramètre d'échelle pour prendre en compte la surdispersion. Sur les modèles 30 (avec la variable *PARTCOM*) et 31 (avec la variable *LIEN5*), on observe que, malgré la modification de l'écart type de chaque variable, toutes les variables restent significatives au seuil de 1%. L'estimation obtenue par l'intermédiaire du modèle présenté ci-dessus est donc robuste. L'interprétation des résultats n'est pas modifiée dans la mesure où les coefficients obtenus sont strictement identiques que ce soit dans le modèle de Poisson ou dans le modèle modifié.

Cette possibilité n'est pas la seule pour prendre en compte l'effet de la surdispersion. Il est aussi possible de ne pas utiliser le modèle de Poisson mais le modèle négatif binomial. Dans la section suivante, nous présentons ce modèle.

4.3 Relâchement des hypothèses du modèle de Poisson

La critique principale formulée à l'égard du modèle de Poisson est l'égalité supposée entre moyenne conditionnelle et variance. L'expression de la variance doit permettre de prendre en compte la dispersion du modèle. Aussi, Abbess et al.(1981) proposent l'expression suivante pour la variance :

$$Var(y_{ij}) = E(\lambda_{ij}) + \theta E^2(\lambda_{ij})$$

Il est donc nécessaire d'utiliser un modèle qui permet d'intégrer une telle hypothèse. Le modèle binomial négatif l'intègre car, dans ce type de modèle, la variance respecte l'hypothèse proposée par Abbess et al. (1981) et diffère de l'espérance. Nous présentons ce modèle dans la sous-section suivante.

4.3.1 Spécification du modèle binomial négatif

Le modèle binomial négatif peut être présenté comme une généralisation du modèle de Poisson en introduisant un effet individuel tel que :

$$\ln \lambda_{ij} = x'_{ij} \beta + \epsilon_{ij} = \lambda_{ij} + \ln u_{ij}$$

avec ϵ_{ij} la perturbation qui reflète soit une erreur de spécification soit l'omission de variables. u_{ij} est supposée suivre une distribution gamma. Greene (2005) propose la normalisation suivante afin que l'estimation du modèle puisse se faire correctement, notamment l'identification de la constante :

$$g(u_{ij}) = \frac{\theta^\theta}{\Gamma(\theta)} e^{-\theta u_{ij}} u_{ij}^{\theta-1}$$

On a alors :

$$\begin{cases} E(y_{ij}|x_{ij}) = \lambda_{ij} \\ V(y_{ij}|x_{ij}) = \lambda_{ij} + \frac{1}{\theta} \lambda_{ij}^2 \end{cases}$$

La fonction de vraisemblance du modèle est la suivante :

$$L(\beta|Y, X) = \prod_{ij} P(y_{ij}|x_{ij}) = \prod_{ij} \frac{\Gamma(\theta + y_{ij})}{\Gamma(1 + y_{ij})\Gamma(\theta)} \left(\frac{\lambda_{ij}}{\lambda_{ij} + \theta} \right)^{y_{ij}} \left(\frac{\theta}{\lambda_{ij} + \theta} \right)^\theta$$

Cette fonction de vraisemblance peut être estimée grâce au maximum de vraisemblance.

Remarque : Si $\theta = 0$ alors le modèle binomial négatif est équivalent au modèle de Poisson.

Les estimations obtenues à partir du modèle binomial négatif sont présentées en annexe A.17. Les modèles 32 et 33 sont équivalents à la différence près que le premier utilise la variable *PARTCOM* pour mesurer la proximité sociale et le second utilise la variable *LIEN5*.

Les estimations des modèles 32 et 33 sont très proches de celles obtenues dans le modèle de Poisson. Les variables sont toutes significatives. De plus, les paramètres de dispersion de chaque modèle sont proches de 1. Le modèle binomial négatif semble donc mieux convenir à nos données.

Le modèle binomial négatif permet de prendre en compte, par l'introduction d'une perturbation, la dispersion entre l'intensité des collaborations des différents binômes de notre échantillon. Si la surdispersion est avérée, il apparaît donc préférable d'utiliser ce modèle. Toutefois, nous montrons dans la section suivante qu'il existe un test permettant de choisir entre ces deux modèles.

4.3.2 Modèle de Poisson ou modèle binomial négatif ?

Si, dans notre cas, la surdispersion de nos observations nous invite à utiliser le modèle binomial négatif plutôt que le modèle de Poisson, il existe un test qui permet de choisir entre ces deux modèles. La principale différence entre les deux modèles concerne l'expression de la variance. Dans le cas du modèle de Poisson (sans correction), la variance vaut λ_{ij} alors que dans le modèle négatif binomial, elle vaut $\lambda_{ij} + \frac{1}{\theta}\lambda_{ij}^2$. Ce test cherche à déterminer quelle est la spécification de la variance la plus adaptée pour nos données ou autrement dit si le paramètre θ est différent de 0.

$$\begin{aligned} H_0 : \theta &= 0 \\ H_A : \theta &> 0 \end{aligned}$$

Afin de déterminer laquelle des deux hypothèses est correcte, nous utilisons alors le test du ratio de vraisemblance (LR) qui est égal à :

$$LR = -2(\log(LL_{Poi}) - \log(LL_{NegBin}))$$

Ce test suit une loi du chi-deux à 1 degré de liberté (Camron et Trivedi, 1998).

$$LR \sim \chi_1^2$$

Dans ce cas, le modèle de Poisson apparaît alors comme le modèle contraint, le modèle négatif binomial comme le modèle non contraint. Le modèle de Poisson est le modèle contraint puisque l'on suppose que l'espérance conditionnelle de l'intensité des collaborations est égale à la variance conditionnelle. Or, nous avons vu en début du chapitre que l'intensité moyenne était inférieure à la variance. Nous calculons ce test pour les modèles avec la variable *PARTCOM* et ceux avec la variable *LIEN5*. Les résultats sont présentés dans le tableau suivant :

Modèles	LR	$Seuil$
Modèle avec <i>Partcom</i>	120,425	5,02
Modèle avec <i>Lien5</i>	122,67	5,02

Tableau 47 : Résultats du test du ratio de vraisemblance.

Dans les deux cas, nous rejetons l'hypothèse nulle. Il apparaît donc que le modèle binomial négatif, étant donnée la dispersion de l'intensité des collaborations, est un modèle plus adapté pour déterminer quels sont les facteurs qui jouent sur l'intensité des collaborations entre deux agents. Ce modèle

permet de prendre en compte la surdispersion dans le modèle.

L'utilisation de données de comptage permet de donner une seconde dimension aux estimations économétriques qui tentent de rendre compte de l'intensité des collaborations entre deux agents. Dans ce chapitre, nous avons mis en avant le fait que les variables qui jouent sur l'existence des collaborations ou non sont aussi des variables pertinentes pour expliquer l'intensité de celles-ci. Nous avons ainsi montré que les variables réseaux et notamment celles qui portent sur la proximité sociale ont un effet déterminant sur l'intensité des collaborations. Deux agents qui ont déjà collaboré dans le 6^{ème} PCRDT seront plus à même de collaborer à nouveau et plusieurs fois ensemble dans le 6^{ème} PCRDT. De plus, les modèles basés sur les données de comptage ont mis en avant, tout comme les modèles logit, l'ambiguïté de l'effet de la proximité géographique sur les choix de collaboration puisque dans le modèle binomial négatif, la variable *Proxgeo* apparaît significative alors que les variables *temps* et *distance* ne le sont toujours pas. Par la suite, il pourrait être intéressant de déterminer s'il n'existe pas des facteurs propres à chaque modèle, par exemple un facteur qui serait pertinent uniquement pour expliquer l'intensité des collaborations.

Les limites de nos estimations sont identiques à celles sur variables dichotomiques. En effet, l'ensemble des informations que nous possédons sur les agents concernés provient uniquement des bases de données PCRDT ce qui induit que nous manquons d'information sur les caractéristiques des agents. De plus, la population étudiée, qui a pour principale caractéristique d'avoir participé aux 5^{ème} PCRDT et 6^{ème} PCRDT, collaborent plus en moyenne que l'ensemble des agents sur ces deux programmes cadres. Pour autant, il est nécessaire que nous possédions des informations sur la participation des agents aux programmes cadres précédents pour mesurer l'effet de proximité sociale notamment. Une dernière remarque que l'on peut faire est que la proximité des résultats entre le modèle logit et le modèle de Poisson en termes d'effets des différents déterminants peut s'expliquer en partie par le fait que le nombre de binômes collaborant plus d'une fois ensemble est faible.

Une possibilité pour améliorer les résultats portant sur l'intensité des collaborations serait d'utiliser un modèle en deux étapes d'Heckman (Heckman, 1979 ; Greene 1981, 2005). Nous avons vu que de nombreuses collaborations potentielles ne se réalisaient pas, d'où une intensité de 0. La solution serait alors de combiner à la fois les modèles logit et les modèles de Poisson. Dans une première étape, nous estimerions la probabilité de collaborer à l'aide d'un modèle logit. Puis, dans une seconde étape, nous conserverions seulement les collaborations d'intensité supérieure ou égale à 1 et nous cherche-

rions quels sont les déterminants qui influencent cette intensité. Ce modèle en deux étapes permettrait ainsi de mettre en avant le fait que l'existence des collaborations et l'intensité des collaborations possèdent des déterminants communs mais aussi des déterminants propres. Compte tenue des données dont nous disposons, il est actuellement difficile de réaliser une telle estimation faute d'un nombre de variables explicatives suffisantes. En effet, l'intérêt principal de ce modèle est de pouvoir associer à chaque étape des variables explicatives qui lui sont propres. Or, nous ne disposons pas actuellement d'un nombre de variables suffisant et qui soient pertinentes pour pouvoir réaliser une telle estimation.

L'utilisation de modèle de comptage a permis d'introduire une nouvelle dimension dans l'analyse économétrique des réseaux sociaux conduite au travers des modèles dichotomiques. Malgré cela, nous nous trouvons toujours dans le cadre de relations bilatérales. L'étude des relations bilatérales a mis en avant un certain nombre de déterminants qui expliquent les choix de collaborations des agents. Pour autant, il pourrait s'avérer intéressant de chercher à expliquer la formation d'autres formes de collaborations que seulement les liens bilatéraux. Nous avons vu notamment dans la première partie de la thèse lors de la présentation des PCRDT que la plupart des instruments avait comme condition la présence d'au moins trois agents de pays différents dans un projet. Pour qu'un projet soit validé, il faut donc qu'il soit composé d'au moins trois agents différents. De la même manière que nous avons cherché à déterminer si les facteurs qui expliquaient l'existence ou non d'une collaboration étaient pertinents pour expliquer l'intensité de ces dernières, il pourrait être intéressant de vérifier si ces mêmes facteurs sont des variables intéressantes pour expliquer la formation de trio (trois agents reliés dans le réseau) au sein des PCRDT. C'est l'objet du chapitre suivant.

Chapitre 5

Dépasser les liens bilatéraux

Dans les deux précédents chapitres, nous avons étudié la formation de liens bilatéraux. Ces derniers sont construits à partir de la co-participation de deux agents à un même projet. Il est réaliste de considérer qu'un agent qui participe à un projet accepte de collaborer avec tous les agents de ce projet. Dans le cas des PCRDT, la plupart des instruments (types de projet) impose la présence au sein du projet d'au moins trois agents de pays différents dont au moins deux issus de l'UE, ce qui implique nécessairement que ces projets comptent au moins trois partenaires. Donc, même si l'existence de liens bilatéraux est une hypothèse réaliste, elle reste une hypothèse forte dans la mesure où un agent accepte un projet avec l'ensemble des agents qui le composent. La présence de certains agents au sein du projet peut s'avérer déterminante dans le choix des autres de participer ou non à ce dernier.

Lorsque l'on étudie les liens bilatéraux, nous nous intéressons seulement aux caractéristiques du binôme sans utiliser les informations que nous possédons sur les projets. En effet, utiliser cette information dans les régressions est impossible puisqu'elle ne serait disponible que pour les agents qui collaborent dans un projet. Une autre possibilité serait de tester toutes les formations possibles de sous-ensembles d'agents parmi les N agents qui participent aux PCRDT. Il existe 2^N sous-ensembles d'agents différents à tester. Même dans un échantillon réduit comme celui utilisé dans cette partie et en limitant la taille des sous-ensembles à la taille maximale des projets observés, le nombre de possibilités est trop important. De plus, la prise en compte dans l'estimation et dans la construction des variables explicatives de sous-ensembles de tailles différentes est difficile, voire impossible.

Il est possible en revanche de relâcher l'hypothèse utilisée jusqu'à présent pour intégrer la condition imposée par l'UE, à savoir la présence d'au moins trois agents dans un projet. Afin de prendre en compte cet élément, nous nous intéressons, dans ce chapitre, non plus à la formation d'un lien bilatéral

entre deux agents mais à la formation de triades, c'est-à-dire à la présence de collaborations entre trois agents. Le fait de nous intéresser à la formation de partenariats à trois permettra de tester plusieurs nouvelles hypothèses. La première des ces hypothèses fait référence au contexte particulier dans lequel se déroule le 6^{ème} PCRDT à savoir le contexte d'adhésion à l'UE de nouveaux États membres. La seconde renvoie à une hypothèse que l'on a partiellement testée dans les deux premiers chapitres à savoir l'impact des connaissances communes sur la probabilité qu'ont deux agents de collaborer.

Comme pour étudier les liens bilatéraux, il est primordial pour l'étude des triades de posséder des informations sur les collaborations antérieures entre les agents. Dans notre cas, il est donc nécessaire que les agents aient participé au 5^{ème} PCRDT et au 6^{ème} PCRDT. C'est pourquoi nous retenons le même échantillon que pour les deux chapitres précédents à savoir les 114 agents qui ont participé aux deux programmes cadres.

Le chapitre est présenté de la manière suivante. La section 5.1 présente plus précisément la notion de triade et son impact sur le modèle et la construction des variables. La section 5.2 adapte les hypothèses déjà présentées dans les deux chapitres précédents aux triades. Enfin, la section 5.3 met en avant de nouvelles hypothèses que l'étude des triades permet de tester.

5.1 Approche par les triades

5.1.1 Définition d'une triade

Dans les réseaux sociaux, une triade renvoie à un sous-ensemble de trois agents et aux possibles relations entre eux. Dans la suite de ce chapitre, nous utiliserons le terme triade dans un cas spécifique : celui où, dans un sous-ensemble de trois agents, chaque agent possède un lien avec les deux autres. De cette manière, les trois agents forment une clique. Cette définition plus stricte a pour but de faciliter la lecture du chapitre.

Dans le cadre des PCRDT, la notion de triade s'applique différemment puisque nous ne connaissons pas les liens que possèdent les agents mais les projets auxquels ils participent. Afin de pouvoir construire les réseaux sociaux propres aux PCRDT, nous avons supposé que tous les agents qui participent à un projet possèdent un lien entre eux. De cette manière, un projet composé de n agents forme une clique de même taille.

Lorsqu'on observe les relations entre trois agents, si ces trois agents possèdent au moins un projet en commun alors ils forment nécessairement une triade. Dans le cas des PCRDT, si la plupart des triades sont obtenues de

cette manière, ce n'est pas la seule possibilité pour que trois agents forment une triade. En effet, il est aussi possible d'obtenir de telle triade à partir de trois projets. Dans chaque projet, on retrouve seulement deux agents par projet.

Remarque : Il aurait été possible de ne s'intéresser qu'à la première forme de triade mais dans le cadre de premières estimations, il nous a semblé intéressant de retenir les deux formes possibles. Dans un premier temps, l'objectif est de vérifier s'il existe des sous-ensembles de trois agents qui possèdent des liens entre eux. Dans un second temps, il aurait pu être intéressant de distinguer les différentes formes de triade en supposant que certains déterminants s'appliquent plus à une forme de triade qu'à une autre.

Il est donc plus simple de définir les triades à partir des liens. Ainsi, une triade est définie de la façon suivante :

$$y_{ijk} = \begin{cases} 1 & \text{si l'agent } i, \text{ l'agent } j \text{ et l'agent } k \text{ collaborent les uns avec les autres,} \\ & i, j \text{ et } k \in N \text{ et } i, j, k \text{ distincts} \\ 0 & \text{sinon} \end{cases}$$

De manière plus formelle, on a :

$$y_{ijk} = \begin{cases} 1 & g_{ij} = 1, g_{ik} = 1 \text{ et } g_{jk} = 1 \\ 0 & \text{sinon} \end{cases}$$

Dans la définition ci-dessus, nous sommes toujours dans le cas de liens symétriques. Il est donc équivalent de parler d'un lien entre i et j et entre j et i quels que soient i et j appartenant à N .

Dans ce chapitre, nous nous intéressons à l'existence ou non de triades. L'intensité des liens n'est donc pas prise en compte. De fait, le modèle utilisé pour tester les déterminants qui incitent trois agents à former une triade est le modèle logit. Contrairement au chapitre 3, chaque variable ne sera pas associée à un binôme d'agents mais à un sous-ensemble de trois agents. Si cette différence ne modifie pas le choix du modèle, elle a un impact sur la construction des variables explicatives. C'est l'objet de la sous-section suivante.

5.1.2 Construction des variables explicatives propres aux triades

Nous avons vu dans le chapitre 3 qu'une des difficultés de l'étude des liens bilatéraux tenait dans la construction des variables explicatives, plus particulièrement lorsque celles-ci ne sont pas dichotomiques. Le problème est identique lorsque l'on étudie les triades. L'introduction de variables explicatives individuelles est impossible car l'indexation de ces variables compte dans l'estimation. Afin de contourner cette difficulté, nous avons alors utilisé la méthode de construction de variables utilisées par Fafchamps, van der Leij et Goyal (2010). Si cette méthode est adaptée à l'étude des relations bilatérales, elle demande à être repensée dans le cas des triades.

Comme pour l'étude des relations bilatérales, à partir des données individuelles sur les agents, nous construisons deux types de variables : un qui rend compte du niveau moyen des trois agents et un autre qui prend en compte la différence entre les trois agents. Si, dans le premier cas, la construction est relativement similaire mis à part le passage de deux à trois éléments, dans le second cas nous verrons qu'il existe plusieurs possibilités. Pour trois agents i , j et k appartenant à l'ensemble des agents N , les variables réseaux se définissent de la manière suivante :

1. La première variable est égale à la moyenne des valeurs de chaque agent :

$$\bar{C} = \frac{C_i + C_j + C_k}{3}$$

Si le signe du paramètre associé à \bar{C} dans l'estimation est positif alors les trois agents avec de forts indicateurs de centralité sont attractifs les uns pour les autres. Au contraire, si le paramètre associé à \bar{C} est négatif alors des agents centraux n'ont pas intérêt à collaborer ensemble. Nous parlons d'agents centraux car ce sont les agents pour lesquels l'impact de \bar{C} sur la probabilité de collaborer est le plus significatif.

2. La deuxième variable est le maximum de la différence de valeurs entre deux des trois agents en valeur absolue (l'important ici étant l'écart entre les valeurs, le recours à la valeur absolue évite tout problème lié au signe) :

$$\Delta C = \text{Max}|C_a - C_b|$$

avec $a \in \{i, j, k\}$ et $b \in \{i, j, k\}$, $a \neq b$. Si le coefficient du paramètre associé à ΔC dans l'estimation est positif alors plus l'écart maximal

entre deux des trois agents en termes de centralité sera important plus la probabilité qu'ils collaborent sera forte. A l'inverse, si le coefficient est négatif alors plus l'écart maximal entre deux des trois agents sera important en termes de centralité moins la probabilité qu'ils collaborent sera forte.

Dans ce chapitre, nous avons choisi de prendre en compte la différence maximale entre deux des trois agents du binôme. De cette manière, nous mettons en avant le plus grand écart au sein du binôme. Ce choix est celui qui permet par la suite de tester au mieux les différentes hypothèses du modèle puisque cette manière de calculer est celle qui reflète le plus la disparité au sein du binôme.

La variable ΔC aurait cependant pu être calculée de deux autres manières :

- Par une moyenne des valeurs absolues des différences deux à deux :

$$\Delta C' = \frac{|C_i - C_j| + |C_i - C_k| + |C_j - C_k|}{3}$$

- Grâce à la distance euclidienne :

$$\Delta C'' = (|C_i - C_j|^2 + |C_i - C_k|^2 + |C_j - C_k|^2)^{\frac{1}{2}}$$

Afin d'illustrer notre choix de ne pas avoir recours à ces indicateurs finalement, prenons le cas simple suivant. Soit une variable fictive C et 6 agents regroupés en deux sous-ensembles de trois agents. Les valeurs sont données dans le tableau suivant :

	Sous-ensemble A			Sous-ensemble B		
Agents	1	2	3	4	5	6
Valeur de C	1	3	5	3	3	3

Les résultats seraient alors les suivants :

Sous-ensemble	A	B
ΔC	4	0
$\Delta C'$	$\frac{8}{3}$	0
$\Delta C''$	$\sqrt{24}$	0

Dans le tableau ci-dessus, on observe que les trois indicateurs rendent compte convenablement des écarts entre agents. Lorsqu'il n'existe aucun écart, les trois indicateurs renvoient la même réponse. Par contre, lorsque les trois agents possèdent des indicateurs aux valeurs différentes, le second indicateur

($\Delta C'$) a tendance à lisser les écarts. En effet, dans la mesure où il prend en compte tous les écarts deux à deux, cet indicateur est toujours plus faible. Le dernier indicateur ($\Delta C''$) au contraire accentue les écarts.

Dans notre cas, il est important de connaître l'écart maximal au sein du sous-ensemble de trois agents. Par exemple, lorsque l'on s'intéresse aux moyens en R&D, l'écart qui est le plus pertinent pour savoir si trois agents décident de collaborer est celui des deux extrêmes. En effet, l'objectif ce type de variable est de savoir si un agent a tendance à collaborer avec des agents qui lui ressemblent ou au contraire recherche une certaine diversité. Dans le cas des triades, en prenant les deux extrêmes, on prend ainsi la plus grande différence qu'il peut exister au sein du sous-ensemble de trois agents.

Comme pour l'étude des relations bilatérales, l'ordre d'indexation n'intervient plus dans le modèle. Dans la section suivante, nous cherchons à vérifier si les résultats obtenus dans le cas de l'étude des relations bilatérales s'appliquent aussi au cas des triades.

5.2 Hypothèses du modèle et variables explicatives

Dans la section 5.2.1, nous rappelons ces hypothèses sur lesquelles est fondé le modèle en les adaptant à l'étude des triades. Dans la section 5.2.2, nous présentons les résultats obtenus par les premières estimations sur l'étude des triades.

5.2.1 Les déterminants traditionnels

Dans les paragraphes suivants, nous adaptons les différentes hypothèses à l'étude des triades. Pour chacune d'entre elles, nous rappelons les variables utilisées pour les traiter et nous présentons les modifications apportées à ces variables dans le but de les adapter à l'étude des triades.

Les effets réseaux

La première hypothèse concerne l'impact de la centralité des agents dans le réseau antérieur sur le réseau actuel. Elle suppose qu'un agent central dans un réseau est un agent attractif dans le cadre de collaborations futures.

L'hypothèse correspondante dans le cas des triades est :

Hypothèse 1a' : La probabilité que trois agents collaborent est fonction de la centralité des agents au sein du réseau formé par le 5^{ème} PCRDT.

Comme pour l'étude des relations bilatérales, nous retenons le degré renforcé comme mesure de la centralité des agents au sein du réseau du 5^{ème} PCRDT. Les variables sont construites à partir de variables continues. Les variables \overline{DR}^T et ΔDR^T sont élaborées comme présenté à la sous-section 5.1.2.

La deuxième hypothèse cherche à tester s'il existe un effet de proximité sociale qui influence les décisions des agents au moment de collaborer.

Hypothèse 1b' : La présence à une période antérieure d'un lien (de liens) entre les trois agents et/ou la proximité au sein de ce réseau entre ces trois agents favorise(nt) les collaborations futures.

La variable renvoie au nombre moyen de partenaires communs entre les trois agents. Ce n'est pas une variable dichotomique. Elle se calcule de la manière suivante :

$$\overline{PARTCOM}^T = \frac{PARTCOM_{ij} + PARTCOM_{jk} + PARTCOM_{ik}}{3}$$

Cette variable permet d'estimer le rôle joué par la proximité sociale dans la formation des triades. Cette proximité est ici mesurée au travers des partenaires communs que peuvent posséder les agents dans le réseau précédent. Ces partenaires communs sont susceptibles d'introduire les agents les uns auprès des autres et ainsi favoriser les collaborations futures. Dans la section suivante, nous montrerons que l'étude des triades permet d'introduire d'une autre manière le rôle de la proximité sociale dans le modèle. Avant cela, nous poursuivons, dans le paragraphe suivant, l'adaptation des indicateurs à l'étude des triades.

Les moyens en R&D

La seconde hypothèse concernait les moyens en R&D dont disposent les agents et qui sont susceptibles d'affecter les collaborations. Comme pour l'étude des relations bilatérales, les moyens en R&D de chaque agent sont calculés comme le coût moyen supporté par les agents dans les projets auxquels ils participent. La somme des coûts auraient été un indicateur intéressant mais elle serait fortement corrélée avec les indicateurs de centralité. De plus,

elle n'aurait pas reflété correctement les moyens dont dispose un agent car les PCRDT ne sont pas forcément l'unique type de projet collaboratif pour lequel un agent effectue des dépenses en R&D (en France, il y a les pôles de compétitivité par exemple). L'hypothèse appliquée à l'étude des triades est la suivante :

Hypothèse 2' : La probabilité que trois agents collaborent est fonction des moyens dont chacun dispose.

Les variables en R&D associées à cette hypothèse sont calculées comme présentée à la sous-section 5.1.2. Les deux variables obtenues sont : \overline{COUT}^T et $\Delta COUT^T$. Comme dans le cas de l'étude des relations bilatérales, un agent disposant de moyens importants est supposé être un partenaire privilégié. De la même manière, une différence importante de moyens entre deux agents peut être un élément défavorable pour une collaboration.

La proximité géographique

Dans le cas des liens bilatéraux, nous avons vu que l'effet de la proximité géographique était ambigu. Ainsi, la distance ou le temps de parcours n'étaient pas significatifs alors que la variable *Proxgeo* qui prend en compte le fait que deux agents soient localisés dans le même pays ou dans des pays frontaliers était significative. Il est alors intéressant d'observer si dans le cas des triades, les résultats obtenus sont identiques.

Hypothèse 3' : La probabilité que trois agents collaborent est d'autant plus forte qu'ils sont géographiquement proches les uns des autres.

Les variables associées à cette hypothèse sont similaires à celles utilisées dans le cas des liens bilatéraux.

Les variables associées à la distance et au temps de parcours sont calculées comme présentées dans la sous-section 5.1.2. Elle sont notées :

- \overline{DIST}^T et $\Delta DIST^T$ pour la distance ;
- \overline{TEMP}^T et $\Delta TEMP^T$ pour le temps de parcours.

Ces variables sont complétées par des variables dichotomiques qui prennent en compte le fait que les agents peuvent être localisés dans le même pays ou dans des pays frontaliers. Elles sont construites de la manière suivante :

$$Proxgeo3^T = \begin{cases} 1 & \text{si chaque agent est situé dans le même pays} \\ & \text{ou dans un pays frontalier aux deux autres} \\ 0 & \text{sinon} \end{cases}$$

et

$$Proxgeo2^T = \begin{cases} 1 & \text{si seulement deux agents respectent les conditions} \\ & \text{énoncées ci-dessus} \\ 0 & \text{sinon} \end{cases}$$

La dernière hypothèse est présentée dans le paragraphe suivant.

La prise en compte de l'espace économique

Dans le cas des liens bilatéraux, nous avons observé que le fait que deux agents appartiennent au cœur (Allemagne, Espagne, France, Grande-Bretagne, Italie) a une influence négative sur la probabilité de collaborer. Il est intéressant de tester cette hypothèse dans le cas des triades.

Hypothèse 4' : La probabilité que trois agents collaborent est influencée par l'appartenance des trois agents ou de deux agents à un sous-ensemble de pays fortement participatifs appelé le cœur.

Les variables relatives à cette hypothèse sont les suivantes :

$$Core - UE3^T = \begin{cases} 1 & \text{si les trois agents sont localisés} \\ & \text{dans les pays du cœur} \\ 0 & \text{sinon} \end{cases}$$

ou

$$Core - UE2^T = \begin{cases} 1 & \text{si exactement deux agents sont localisés} \\ & \text{dans les pays du cœur} \\ 0 & \text{sinon} \end{cases}$$

ou

$$Core - UE1^T = \begin{cases} 1 & \text{si exactement un agent est localisé} \\ & \text{dans un des pays du cœur} \\ 0 & \text{sinon} \end{cases}$$

Par la suite, nous verrons que le cadre des triades permet d'introduire une hypothèse liée au contexte particulier des PCRDT.

Dans la sous-section suivante, nous présentons les premières estimations sur la formation de triades, estimations qui testent les différentes hypothèses présentées ci-dessus et qui s'appliquent donc au modèle tel qu'il avait été défini pour les relations bilatérales.

5.2.2 Estimations des déterminants traditionnels

Dans cette sous-section, nous présentons les premières estimations liées aux triades. Dans le paragraphe suivant, nous montrons que le faible nombre de triades réalisées au sein des 114 agents induit une légère modification de l'estimation.

Un nombre de triade trop faible

Le tableau ci-dessous présente le nombre de triades réalisées parmi les 114 agents.

	FP5 Nanotechnologies	FP6 Nanotechnologies
Nombre de triades	790	7111
Nombre de triades potentielles	240464	240464
Pourcentage réalisé	0,3%	3%

Tableau 48 : Présence des triades dans le domaine des nanotechnologies pour le 5^{ème} PCRDT et le 6^{ème} PCRDT.

Sur ce tableau, on observe que le nombre de triades formées dans le 6^{ème} PCRDT est très faible. Afin d'obtenir un échantillon plus équilibré, dans nos estimations, nous allons suréchantillonner les observations $y_{ijk} = 1$. Ce suréchantillonnage introduit des modifications dans l'estimation.

Pour réaliser l'estimation, nous utilisons l'estimateur du maximum de vraisemblance d'un échantillon endogène pondéré développé par Manski et

Lerman (1977) et repris par Greene (2005). Pour pouvoir utiliser cet estimateur, il faut connaître la proportion de 1 et de 0 dans la population originale, c'est-à-dire la proportion de 1 et de 0 parmi les 240464 triades potentielles. Ces proportions sont notées respectivement w_1 et w_0 . De plus, on note p_1 et p_0 les proportions de 1 et de 0 dans le nouvel échantillon où le nombre de $y_{ijk} = 1$ est suréchantillonné.

Pour pouvoir obtenir les coefficients des paramètres, il faut alors maximiser la log-vraisemblance pondérée suivante :

$$\ln L = \sum_{T=1}^t w_T \ln F(q_T \beta' x_T)$$

avec T les triades, t le nombre de sous-ensembles de trois agents dans N , $w_T = y_T(w_1/p_1) + (1 - y_T)(w_0/p_0)$ et $q_T = 2y_T - 1$. On remarque ici que w_T ne peut prendre que deux valeurs.

Pour obtenir les estimateurs, une correction finale est nécessaire. En effet, l'estimateur approprié de la fonction de covariance est $H^{-1}BH^{-1}$ où H est le hessien et B est un autre estimateur de la matrice de variance de Hall et Hausman. Ces deux estimateurs sont présentés ci-dessous.

$$H = \frac{\partial^2 \ln L}{\partial \beta \partial \beta'} = - \sum_T \Lambda_T (1 - \Lambda_T) x_T x_T'$$

$$B = \sum_T (y_T - \Lambda_T)^2 x_T x_T'$$

Dans le paragraphe suivant, nous présentons les premiers résultats sur la formation de triades.

Premières estimations sur l'existence des triades

L'ensemble des estimations est présentée en annexe A.18. (modèles 34 à 43). Pour construire le modèle, nous suivons une méthode similaire à celle utilisée pour l'étude des relations bilatérales à savoir que nous commençons par estimer un modèle réduit avec peu de variables qui servira de base à l'estimation, puis nous intégrons de nouvelles variables au fur et à mesure pour tester les hypothèses. De plus, nous allons vérifier qu'en dehors du suréchantillonnage réalisé, le nouvel échantillon possède des caractéristiques relativement similaires pour les autres variables.

Le modèle 34 est constitué des variables \overline{DR}^T , ΔDR^T , \overline{COUT}^T , $\Delta COUT^T$ et $PARTCOM^T$. La première remarque que l'on peut faire est que contrairement à l'étude des relations bilatérales, les variables relatives aux coûts sont non significatives. Il est possible que les moyens en R&D soient un déterminant important lorsqu'un agent s'investit dans une collaboration bilatérale. Par contre lorsque trois agents s'associent simultanément, ils peuvent rechercher à remplir d'autres conditions telles que celles imposées par l'UE pour que le projet soit financé. Dans ce cas, les variables liées aux moyens en R&D n'expliquent pas la formation des triades. Nous verrons que ce premier résultat est valable pour l'ensemble des estimations portant sur les triades. Dans ce modèle, on observe que les variables relatives à la centralité des agents ont toutes les deux un effet significatif et influencent la formation des triades comme elles influençaient la formation des liens bilatéraux. En effet, la variable \overline{DR}^T a une influence positive sur la probabilité que trois agents forment une triade tout comme la variable ΔDR^T . Ceci laisse supposer que le réseau a bien la forme d'étoiles inter-reliées avec des collaborations entre agents centraux et, dans le même temps, des collaborations entre un agent central et un agent moins impliqué dans le réseau. Il est possible de retrouver ce schéma au sein même de la triade avec deux agents centraux qui collaborent ensemble avec un troisième agent qui l'est moins. Enfin, la variable $PARTCOM^T$, qui correspond au nombre moyen de partenaires commun entre les trois agents a un effet significatif et positif sur la probabilité de formation de la triade. La proximité sociale influence donc bien la probabilité de formation des triades. Nous verrons à la section suivante si le phénomène de "triadic closure" est vérifié. Par la suite, quel que soit le modèle estimé, ces 5 variables conservent leur significativité et leur effet sur la probabilité de formation d'une triade.

Les modèles 35 à 39 introduisent les variables qui permettent de prendre en compte la proximité géographique.

Les modèles 35 à 37 intègrent la proximité géographique en introduisant les variables binaires $PROXGEO3^T$ et $PROXGEO2^T$ de manière séparée (modèles 35 et 36) puis simultanément (modèle 37). On observe que la variable $PROXGEO3^T$ a un effet significatif et positif sur la formation des triades dans les deux modèles où elle apparaît. En revanche, la variable $PROXGEO2^T$ a un effet non significatif à chaque fois. Il apparaît donc que le fait que les trois agents soient localisés dans le même pays ou dans des pays frontaliers augmente la probabilité de formation des triades. Dans le cas de la formation des liens bilatéraux, nous avons montré que la variable $PROXGEO$ avait un effet similaire. Dans le cas des triades, c'est la proximité des trois agents qui importent. Si seulement deux agents sont proches l'un

de l'autre ce n'est pas suffisant. Il s'agit donc d'une condition de proximité forte.

Dans les modèles 38 et 39, nous introduisons la proximité géographique par l'intermédiaire des variables relatives au temps et à la distance (par route dans les deux cas) qui séparent les agents. Dans l'étude des relations bilatérales, ces variables s'étaient révélées non significatives. Dans les deux modèles, seules les variables exprimant la distance moyenne séparant les trois agents, \overline{TEMPS}^T et \overline{DIST}^T ont un effet significatif et négatif sur la probabilité que trois agents forment une triade. Ceci signifie que plus le temps de parcours ou la distance entre les trois agents sont élevés moins la probabilité qu'une triade se forme est élevée. Ce résultat est cohérent avec celui obtenu à partir de la variable $PROXGEO3^T$. Une raison qui peut expliquer que ces variables ont un effet significatif dans le cas des triades et non dans celui des liens bilatéraux est le fait que le calcul de la moyenne sur trois distances deux à deux permet de lisser les valeurs de ces variables et ainsi éviter des valeurs élevées comme il était possible d'en avoir dans l'étude des relations bilatérales.

La dernière hypothèse que nous souhaitons tester dans cette section est l'importance que peut avoir la localisation des agents dans les pays du cœur (Allemagne, Espagne, France, Italie et Royaume-Uni) qui sont, pour rappel, les pays les plus participatifs au sein des PCRDT. Pour cela, nous avons construit trois variables $CORE-UE3^T$, $CORE-UE2^T$ et $CORE-UE1^T$. Ces trois variables sont introduites séparément (modèles 40 à 42) puis simultanément (Modèle 43). Les variables $CORE-UE3^T$ et $CORE-UE2^T$ ont toutes les deux un effet significatif et négatif, qu'elles soient introduites séparément ou simultanément. La variable $CORE-UE1^T$ a un effet significatif et positif lorsqu'elle est introduite seule et a un effet non significatif lorsqu'on l'introduit simultanément avec les deux autres variables. On observe donc que compter trois ou deux agents localisés dans le cœur dans un sous-ensemble de trois agents affecte négativement la probabilité que ces trois agents forment une triade. Par contre, on observe que compter au moins un agent du cœur a un effet positif.

Au travers de ces estimations, la manière dont se forme les triades semble se préciser. La principale différence avec l'étude des relations bilatérales réalisées au chapitre 3 est l'absence de rôle joué par les moyens en R&D dans la formation des triades. En outre, on observe que l'effet de la proximité géographique est plus claire dans le contexte des triades. La section suivante a pour but de compléter les estimations réalisées dans cette section en testant

deux nouvelles hypothèses.

5.3 La prise en compte du contexte particulier des PCRDT

Dans la section précédente, nous avons cherché à vérifier si les hypothèses mises en avant dans l'étude des liens bilatéraux étaient aussi valables dans le cadre des triades et si les déterminants avaient les mêmes effets. A partir du modèle de base obtenu dans la section précédente, nous introduisons deux nouvelles hypothèses que le cadre spécifique aux triades nous permet de tester. La première est relative à la proximité sociale et la seconde renvoie aux conditions particulières imposées par l'UE dans les projets des PCRDT.

5.3.1 “Triadic closure” : un rôle accru de la proximité sociale

L'étude des triades permet de tester la “triadic closure”. La “triadic closure” est un concept qui a été suggéré par un sociologue allemand Georg Simmel au début du $XX^{\text{ème}}$ siècle. Puis, le concept de “triadic closure” a été repris par Mark Granovetter dans son article de 1973 “The Strength of the Weak Ties”.

Définition et hypothèse

La “triadic closure” caractérise les relations entre trois agents. Cette notion est utilisée dans l'analyse des réseaux sociaux. La définition originale est plus précise que celle que nous retenons. Pour trois agents i , j et k , s'il existe un lien fort entre i et j et entre j et k alors il existe un lien faible ou un lien fort¹ entre i et k .

La différence entre liens forts et liens faibles n'est pas l'objet de notre étude bien que comme vu précédemment, l'intensité des liens peut permettre d'adapter ces notions au cas des PCRDT. Dans notre cas, nous reprenons la définition utilisée par Easley et Kleinberg (2010, p44) :

“If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point

¹La notion de liens faibles et de liens forts renvoie à deux types de relations possibles entre deux agents. De manière simple, les liens forts sont caractérisés par des rencontres fréquentes et des échanges approfondis alors que des liens faibles correspondent à des contacts brefs et occasionnels.

in the future.”

Cette définition correspond mieux au cadre des PCRDT et à notre objet d’étude. Par analogie, il est possible d’adapter cette définition aux relations entre participants dans les PCRDT. Dans cette définition, la dimension temporelle est mise en avant. Ainsi, c’est la situation à un instant t qui est susceptible d’influencer la situation en $t + 1$. Dans le cadre de la thèse, cela se traduit par le fait que les collaborations dans le 5^{ème} PCRDT peuvent influencer celles du 6^{ème} PCRDT. La notion de “triadic closure” et l’étude des triades nous offrent donc un cadre idéal pour tester une autre hypothèse liée à la proximité sociale.

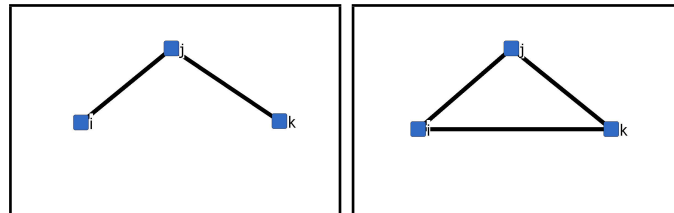
L’hypothèse que nous souhaitons tester est la suivante :

Hypothèse 1c’ : Lorsque l’on s’intéresse à un sous-ensemble de trois agents, si deux de ces agents ne possèdent aucun lien dans le réseau du 5^{ème} PCRDT mais ont un partenaire commun, alors cela augmente la probabilité que les trois agents (les deux agents non liés et le partenaire commun) collaborent à la période suivante.

La variable associée à cette hypothèse est construite de la manière suivante :

$$TC_{ijk} = \begin{cases} 1 & g_{ij} = 1 \quad g_{jk} = 1 \quad g_{ik} = 0 \text{ dans le 5^{ème} PCRDT} \\ 0 & \text{sinon} \end{cases}$$

L’hypothèse ci-dessus est plus stricte que la définition donnée précédemment puisque l’on souhaite que les deux agents non lié qui possèdent un partenaire commun collaborent tout en maintenant leurs liens à la période suivante avec leurs partenaires communs. Nous cherchons à modéliser le cas suivant :



Graphes 17 et 18 : Illustration du phénomène de “triadic closure”.

La situation dans le graphe de gauche correspond aux relations de trois agents dans le 5^{ème} PCRDT. Si cette situation se présente et que l'hypothèse est vérifiée alors on devrait aboutir au graphe de droite dans le 6^{ème} PCRDT. Il existe deux possibilités pour cela : soit les trois agents collaborent dans un même projet soit ils collaborent deux à deux dans trois projets différents.

Dans le cas des réseaux sociaux, plusieurs raisons ont été identifiées pour expliquer un tel phénomène :

- L'opportunité de se rencontrer. A partir du moment où deux agents possèdent une connaissance commune alors la probabilité qu'ils se rencontrent devient plus importante. Dans le cas où l'on étudie des relations amicales, c'est souvent le temps passé avec chacun qui fait que, deux personnes qui ne se connaissent pas, finissent par se rencontrer. Dans le cas des PCRDT, il est possible que ces rencontres se déroulent dans des contextes professionnels du type réunions, séminaires, ... De plus, le partenaire commun peut très bien introduire les deux autres agents l'un auprès de l'autre.
- La confiance. Le fait que deux agents possèdent un partenaire commun favorisera leur rapprochement dans la mesure où s'ils ont confiance dans leur partenaire, il est alors plus facile d'avoir confiance en un partenaire d'un partenaire (Gulati, 1993). De plus, dans un contexte d'incertitude, le fait de pouvoir avoir des informations sur un futur partenaire est un avantage, permet de limiter les risques et d'évaluer, en amont de la collaboration, les bénéfices que l'on peut retirer de cette dernière en fonction des caractéristiques de cet agent.
- La pression. Dans le cas d'une relation amicale, si i est ami avec j et k mais que j et k ne le sont pas alors cela peut être source de stress pour l'agent i . Dans le cas de partenariats au sein des PCRDT, il est délicat pour un agent i de collaborer ensemble avec deux agents qui ne veulent pas collaborer surtout quand les deux collaborations ont lieu dans le même domaine technologique. Dans le cas de relations de collaborations où la confidentialité est importante, il peut être difficile pour l'agent au centre de collaborer sur deux projets dans le même domaine sans ne rien dévoiler.

Les deux premières raisons semblent être celles qui s'adaptent le mieux au cadre des collaborations en R&D au sein des PCRDT.

Dans le paragraphe suivant, nous introduisons la variable TC dans le modèle de référence afin de tester l'hypothèse présentée ci-dessus.

Estimation

Dans le modèle 44, nous avons repris le modèle de base complété par les variables $CORE - UE3^T$ et $PROXGEO3^T$. Les résultats sont présentés dans l'annexe A.18.

On observe que les variables que l'on a présentées précédemment conservent leur significativité. De plus, elles conservent aussi leur effet, positif ou négatif, sur la formation d'une triade. Concernant la variable TC^T qui rend compte du phénomène de "triadic closure", elle a un effet significatif et positif sur la probabilité qu'une triade se forme. Cette estimation permet donc de vérifier empiriquement que lorsque l'on a un sous-ensemble de trois agents i , j et k à une période $t - 1$ tel que i et j soient liés et j et k aussi alors cela influence positivement la formation d'une clique entre ces trois agents à la période suivante.

Nous avons réalisé une seconde estimation en ôtant la variable $PARTCOM^T$ pour introduire seulement la variable TC^T pour prendre en compte le phénomène de proximité sociale. La variable TC^T a toujours un effet significatif et positif sur la formation des triades.

Cette estimation met en avant qu'il existe une continuité entre les deux programmes cadres en termes de collaboration. Nous avons déjà remarqué un phénomène similaire dans l'étude des relations bilatérales avec la variable $LIEN5$. Dans les deux cas, cela montre l'importance des collaborations antérieures pour expliquer les collaborations présentes.

5.3.2 Les agents des nouveaux États membres incontournables au sein des triades ?

Nous avons vu qu'une des spécificités des projets financés par l'UE dans le cadre des PCRDT était d'imposer d'avoir au minimum trois agents localisés dans des pays différents dont au moins trois pays de l'UE des 15 ou des nouveaux États membres. De plus, le 6^{ème} PCRDT se déroule dans un contexte d'adhésion qui peut favoriser la participation des agents localisés dans les nouveaux États membres. Il est intéressant de vérifier de quelle manière les agents localisés dans les nouveaux États membres s'intègrent aux dispositifs.

Hypothèse

Les 114 agents qui composent notre échantillon sont localisés pour la grande majorité dans un des pays de l'UE. Seuls quelques agents sont locali-

sés dans des pays associés comme la Suisse ou la Norvège. Tester la présence de trois agents de l'UE au sein d'une triade aurait donc peu d'intérêt. Par contre, il est intéressant de chercher à étudier de quelle manière les agents localisés dans les nouveaux États membres s'intègrent aux PCRDT. Dans un contexte d'intégration des nouveaux États membres, une collaboration avec un agent localisé dans un de ces pays peut être vue comme une condition favorable au financement. Ainsi deux agents de l'UE des 15 peuvent avoir intérêt à collaborer avec un troisième agent issu des nouveaux États membres. C'est l'hypothèse que nous cherchons à tester.

Hypothèse 4b' : Si dans un sous-ensemble de trois agents, deux sont localisés dans l'UE des 15 et un est localisé dans un nouvel État membre alors la probabilité que les agents collaborent est plus élevée.

La variable qui permet de contrôler cet effet est la suivante :

$$Pays_{ijk} = \begin{cases} 1 & \text{si } i \text{ et } j \text{ sont localisés dans un pays de l'UE des 15} \\ & \text{et } k \text{ dans un nouvel état membre} \\ 0 & \text{sinon} \end{cases}$$

Grâce à cette variable, nous pouvons contrôler si deux agents de l'UE des 15 sont incités à collaborer avec des agents localisés dans des nouveaux états membres.

Pour tester cette hypothèse, nous intégrons la variable définie au modèle de base. C'est l'objet du paragraphe suivant.

Estimation

Dans cette sous-section, nous cherchons à montrer que la manière dont est composé le sous-ensemble de trois agents peut jouer sur la formation d'une triade ou non. Pour cela, nous avons introduit dans le modèle de base la variable $Pays^T$. Dans ce modèle, nous n'introduisons aucune autre variable relative à la proximité géographique ou à l'appartenance au cœur. Les résultats des estimations sont donnés dans le modèle 46 en annexe A.18.

Les variables du modèle de base conservent leur significativité ainsi que leur effet positif ou négatif sur la formation des triades. La variable $Pays$ a un effet significatif et positif. Ce résultat signifie que si le sous-ensemble de trois agents est composé de deux agents localisés dans les pays de l'UE à 15

membres et d'un autre localisé dans un nouvel État membre alors cela favorise la réalisation d'une triade. Ce résultat a pu être influencé par le contexte du 6^{ème} PCRDT qui se déroulait dans un contexte d'adhésion des nouveaux États membres.

Il serait intéressant de vérifier si c'était le cas dans les programmes cadres précédents. Comparer les situations lors de différents programmes cadres avant l'adhésion (5^{ème} PCRDT et avant), pendant l'adhésion (6^{ème} PCRDT) et une fois le processus terminé (7^{ème} PCRDT) permettrait d'observer l'évolution du rôle joué par les agents localisés dans les nouveaux États membres et de déterminer de manière plus claire le rôle joué par les incitations au sein des programmes cadres. Sur les 240464 sous-ensembles de trois agents possibles, 30906 sous-ensembles sont constitués de cette manière. Dans le 5^{ème} PCRDT, sur les 790 triades, 53 étaient constituées de deux agents de l'UE à 15 membres et d'un nouvel État membre soit 6,7% des triades constituées. Dans le 6^{ème} PCRDT, cette proportion augmente considérablement puisque ce sont 16,4% des triades formées qui sont constituées de la sorte. Ces chiffres combinés aux estimations réalisées semblent montrer l'importance des incitations de l'UE dans la constitution des triades.

Nous avons vu au début du chapitre que les triades pouvaient être formées de deux façons : les triades constituaient par la participation de trois agents à un même projet et les triades construites sur la base de collaboration deux à deux entre les trois agents. Il paraît probable que la variable *Pays* aurait eu un effet plus important si l'on avait considéré seulement le premier type de triade.

L'étude de la formation des triades a permis de réaliser des estimations en dehors du cadre traditionnel (étude des relations bilatérales). Ce cadre permet d'une certaine manière d'être plus proche de la réalité des PCRDT puisque certains instruments demandaient un minimum de trois agents. L'étude des triades a surtout permis d'aller plus loin dans l'étude du rôle joué par les collaborations antérieures sur les collaborations présentes. En effet, l'étude de la formation des triades a permis de mettre en avant le rôle joué par la proximité sociale notamment grâce à la vérification de l'hypothèse portant sur la notion de "triadic closure". De plus, le rôle joué par les variables relatives à la centralité des agents est confirmé. Ce chapitre a aussi mis en évidence des différences avec l'étude des relations bilatérales. Les deux différences concernent les variables sur les moyens en R&D et le rôle joué par la proximité géographique. Les variables concernant les moyens en R&D ont un effet non significatif sur la formation de triades. En revanche, l'effet de

la proximité géographique est moins ambigu pour les triades. Les variables $PROXGEO3^T$, \overline{TEMPS} et \overline{DIST} montrent clairement que la proximité entre trois agents favorise leur collaboration. Enfin, le cadre de l'étude des triades a permis de montrer que la présence au sein d'un sous-ensemble de trois agents de deux agents de l'UE à 15 membres et d'un agent localisé dans un nouvel Etat membre favorise la formation d'une triade.

En termes de politique publique, l'UE cherche à favoriser l'intégration au sein de l'EER des agents localisés dans les nouveaux États membres. Pour cela, nous avons vu que l'UE posent certaines contraintes selon les instruments. Au travers de l'étude des triades, nous avons vu que ces contraintes étaient respectées et que le nombre de triades comprenant un agent localisé dans un nouvel État membre avait augmenté de manière considérable entre le 5^{ème} et le 6^{ème} PCRDT. Pour autant, il est difficile de déterminer si ces agents sont recherchés pour leur réelle capacité scientifique ou s'il sont recherchés de manière contrainte afin de respecter les conditions imposées quel que soit leur potentiel scientifique de ces derniers. Nous avons montré antérieurement que le rattrapage des nouveaux membres de l4UE n'avait pas eu lieu et que la participation de ces agents au réseau était faible prise individuellement.

Conclusion

Dans la première partie, nous avons analysé la structure des réseaux établis et, dans cette partie, nous expliquons en partie la manière dont s'est formé le réseau au travers des déterminants qui ont incité les agents à collaborer les uns avec les autres au sein des PCRDT. L'objectif était de montrer que les choix de collaborations dans le 6^{ème} PCRDT étaient influencés à la fois par des déterminants traditionnels (moyens en R&D et proximité géographique) que l'on retrouve dans la littérature sur les collaborations et par d'autres déterminants relatifs au réseau antérieur du 5^{ème} PCRDT. L'économétrie de réseau utilisé dans cette partie nous permet aussi de mieux comprendre certaines conclusions vues dans la partie précédente. Grâce à l'économétrie de réseau, nous avons cherché à expliquer les choix de collaborations de plusieurs manières complémentaires : en étudiant l'existence de collaborations deux à deux dans un cadre de relations bilatérales (chapitre 3), en étudiant l'intensité des collaborations entre deux agents (chapitre 4) et en étudiant la formation de triades (chapitre 5).

Dans le chapitre 3, l'étude des déterminants des choix de collaborations bilatérales a permis de montrer l'importance des collaborations antérieures pour expliquer les collaborations présentes. Nous avons ainsi pu expliquer en partie les réseaux de collaboration du 6^{ème} PCRDT à partir des collaborations du 5^{ème} PCRDT. Au delà des effets de proximité sociale entre agents, le prestige acquis par les agents dans le réseau de collaboration du 5^{ème} PCRDT, mesuré via les indicateurs de réseau, est aussi une variable qui joue sur les collaborations présentes. Concernant les déterminants plus traditionnels, nous avons montré que les moyens en R&D avaient un effet positif sur la probabilité de collaborer. Dans le même temps, la proximité géographique joue un rôle ambigu au sens où toutes les variables utilisées pour tester cette hypothèse ne se sont pas révélées significatives.

Ensuite, nous avons montré que les déterminants des collaborations entre

deux agents n'étaient pas homogènes et pouvaient dépendre des caractéristiques de ces deux agents mais aussi de leurs collaborations passées. Le fait que deux agents aient collaboré à la période précédente semble introduire une relation de confiance qui par la suite suffit à maintenir la collaboration même si nous avons remarqué que la proximité géographique et les moyens en R&D étaient des déterminants pertinents. La proximité géographique favorise alors la poursuite de la collaboration. Dans le cas où les agents n'ont pas collaboré précédemment, le modèle est plus pertinent et met en avant les variables réseaux (le prestige de l'agent et la présence de partenaires communs dans le 5^{ème} PCRDT) qui soulignent une nouvelle fois le rôle de la proximité sociale et des réseaux de collaborations antérieures. La présence ou non d'une collaboration à la période antérieure n'est pas la seule cause d'une hétérogénéité de comportements. Nous avons aussi montré que la diversité des agents était source de comportements hétérogènes. Les modèles pertinents ne sont pas les mêmes pour les entreprises ou pour les centres de recherche.

Dans les deux chapitres suivants, nous sommes sortis du cadre traditionnel, à savoir l'étude de l'existence de collaborations bilatérales, pour étudier l'intensité des collaborations bilatérales d'une part et la formation de triades d'autre part. Dans les deux cas, nous avons utilisé une méthode identique puisque nous avons à chaque fois cherché à vérifier si les déterminants des relations bilatérales s'appliquaient aussi à l'intensité des liens et à la formation de triades.

Lorsque nous avons étudié l'intensité des collaborations entre deux agents, nous avons montré que les déterminants utilisés pour étudier l'existence de relations bilatérales permettait aussi d'expliquer l'intensité des collaborations bilatérales. Les variables significatives dans le premier cas se sont aussi révélées significatives pour expliquer l'intensité des collaborations et ont conservé leur effet (positifs ou négatif). Ces estimations ont une nouvelle fois mis en avant le rôle prépondérant des collaborations réalisées dans le 5^{ème} PCRDT pour expliquer l'intensité des collaborations dans le 6^{ème} PCRDT. Toutefois, nous avons remarqué que le nombre de 0 et le nombre de 1 pour la variable expliquée étaient très importants ce qui laisse penser qu'une autre forme de modèle pourrait être utilisé afin de prendre en compte ce phénomène. Un modèle en deux étapes d'Heckman serait d'une certaine façon une synthèse entre les deux premiers chapitres de la partie 2. Dans une première étape, nous étudierions l'existence ou non d'une collaboration entre deux agents comme réalisée dans le chapitre 3. Puis, dans une seconde étape, nous étudierions l'intensité des liens seulement pour les collaborations réalisées et non plus sur l'ensemble des collaborations potentielles. Ce modèle en deux étapes permettrait de prendre en compte le fait que ces deux étapes possèdent des

déterminants communs mais qu'il existe aussi des déterminants propres à l'existence des collaborations et d'autres qui expliquent seulement l'intensité de la collaboration.

Enfin, la dernière approche est plus originale car, contrairement à la plupart des études sur les réseaux de collaboration, elle s'applique à étudier la formation de triades, c'est-à-dire la réalisation d'une clique entre trois agents au sein du réseau. L'étude de la formation de triades a été retenue afin de chercher à être au plus près du cadre imposé par l'UE. En effet, une partie des instruments impose que le projet soit soutenu par au moins trois agents. On peut supposer, à juste titre, que chaque agent qui accepte de participer à un projet accepte de collaborer avec l'ensemble des agents qui y participent également. Ainsi, il existe un bénéfice pour l'agent de collaborer avec chacun des agents pris un à un et de ce fait une approche par les liens bilatéraux est justifiée. L'étude de la formation de triades ne remet pas en cause cela, elle apporte juste un complément d'explication sur la stratégie que peut adopter un sous-ensemble de trois agents au moment de constituer leur projet pour le soumettre à l'UE. L'approche par les triades s'inscrit donc en complément de celles sur les relations bilatérales. Nous avons montré que cette approche permettait de mettre en avant deux principales conclusions. La première c'est que le réseau des collaborations de la période précédente joue un rôle important dans l'établissement des collaborations au sein du réseau de la période suivante. En effet, nous avons montré que le phénomène de "Triadic closure" était présent au sein des réseaux. Les agents ont donc tendance à rapprocher pour une collaboration future leurs connaissances au sein du réseau afin d'établir un nouveau projet. La deuxième conclusion concerne la composition des triades. Nous avons montré que le contexte et les incitations de l'UE pour faire participer des agents localisés dans les nouveaux États membres jouaient sur les choix des partenaires. Ainsi le nombre de triades contenant exactement un agent localisé dans un nouvel État membre et deux agents localisés dans l'UE à 15 membres a considérablement augmenté entre le 5^{ème} PCRDT et le 6^{ème} PCRDT.

Le choix de s'intéresser uniquement aux agents qui ont participé aux deux programmes cadres limite la portée des conclusions faites de cette partie puisque ces conclusions ne s'appliquent qu'aux agents qui possèdent cette caractéristique. Ce choix est justifié dans la mesure où par cette caractéristique les agents se distinguent dans les PCRDT et que leurs choix sont guidés en partie par leurs collaborations passées. De ce fait, il est difficile de supposer que les déterminants des choix de collaborations sont identiques pour l'ensemble des agents qu'ils aient ou non participé au précédent programme cadre. Afin de compléter nos estimations, il pourrait être opportun

de montrer cette différence. En effet, selon que les agents aient participé au précédent programme cadre ou non, certains déterminants comme les moyens en R&D ou la proximité géographique peuvent avoir plus ou moins d'importance. Dans le cas où les agents participent pour la première fois aux PCRDT, on peut supposer que ces déterminants jouent encore plus favorablement en faveur des collaborations. On peut supposer que le prestige des agents mesurés par les indicateurs de centralité peut aider un agent à se décider. Il serait donc intéressant de montrer que les déterminants des collaborations diffèrent entre les binômes selon qu'aucun, seulement un ou les deux agents aient déjà participé au programme cadre. Nous pourrions utiliser un test de changements structurel pour montrer cette hétérogénéité, puis, dans un second temps, chercher les déterminants propres à chaque composition de binôme. L'autre limite du modèle tient au fait que nous nous sommes intéressés qu'à un seul sous-domaine technologique à savoir celui des nanotechnologies et seulement à deux PCRDT limitant ainsi les champs sectoriels et temporels. De ce fait, les conclusions que l'on a pu retirer des estimations peuvent n'être valables que pour ce sous-domaine.

Dans les travaux futurs, un recours à des données plus larges permettrait d'améliorer la portée des travaux de la thèse. Cette extension peut prendre plusieurs formes selon que l'on recourt à des données supplémentaires sur les PCRDT ou en utilisant des données extérieures aux PCRDT.

Une première possibilité serait d'élargir les domaines prises en compte soit à un domaine technologique soit à une thématique soit à l'ensemble du 6^{ème} PCRDT. Cet élargissement demandera un énorme travail en termes d'identification des agents. De plus, il faudra déterminer dans quelle mesure il existe une continuité entre le 5^{ème} PCRDT et le 6^{ème} PCRDT en termes de thématiques abordées. De cette manière, il sera aussi possible de déterminer s'il existe des particularités sur les déterminants des choix de collaborations en fonction des thématiques.

En plus d'élargir les domaines technologiques retenus, disposer de données sur un plus grand nombre de périodes, donc sur un plus grand nombre de PCRDT, permettrait d'avoir recours à un modèle de panel qui autoriserait une meilleure prise en compte des variables retardées et plus particulièrement de la variable expliquée retardée. L'utilisation de données de panel permettrait aussi de suivre la trajectoire des agents d'un programme cadre à l'autre. Ces données permettraient surtout de mettre en avant dans quelle mesure deux agents ont tendance à répéter leur collaborations dans le temps et s'il existe un phénomène d'usure qui fait qu'au bout d'un certain temps les deux agents peuvent ne plus avoir d'intérêt à collaborer.

Dans le cas des PCRDT, nous pourrions ainsi vérifier dans quelle me-

sure les interactions répétées entre agents leur sont favorables en termes de performance économique. Goerzen (2007) s'est intéressé à l'impact des collaborations répétées sur les performances des entreprises japonaises. Pour cela, il s'est basé sur une étude menée en 1999 sur 13529 filiales de 580 grandes multinationales japonaises. Il a notamment mis en avant une relation négative entre la répétition des collaborations et la performance des entreprises. Cette relation est basée sur le postulat qu'une entreprise doit sans cesse mettre à jour ses connaissances et ses compétences et créer de nouveaux savoirs à un coût le moins élevé possible de la même manière que la performance d'une entreprise et la vitalité de l'entreprise dépendent de sa capacité à produire de nouveaux produits et process. Goerzen a ainsi montré que les firmes qui réalisent les plus mauvaises performances étaient aussi celles qui créent le plus grand nombre de joint ventures. Enfin, Goerzen a montré que l'effet négatif liant collaborations répétées dans le temps et performance économique est plus important dans les environnements où l'incertitude technologique est plus importante. Notons que cette étude demanderait des données de panel avec un nombre de périodes suffisantes qui demanderait certainement de devoir découper chaque PCRDT en sous-périodes.

Enfin, comme dans la partie 1, l'une des possibilités pour améliorer ce travail est de chercher à améliorer la construction des réseaux. Actuellement, le travail d'estimation repose toujours sur l'hypothèse que tous les agents d'un projet ont un lien entre eux. Si l'on pouvait déterminer les liens réels au sein de chaque projet, on pourrait ainsi étudier avec plus de précisions les déterminants des choix de collaborations. La solution la plus envisageable est d'utiliser les données sur les coûts supportés et les financements demandés au sein des PCRDT pour pouvoir affiner la construction du réseau. Cette méthode est encore difficilement applicable puisque ces données ne sont pas fournies pour l'ensemble des agents. De plus, cette méthode pose un certain nombre de problèmes méthodologiques notamment la détermination de seuil pour décider si une participation est jugée suffisante, seuil pouvant varier en fonction du type des agents.

La seconde possibilité d'extension de notre travail serait de recourir à des données extérieures aux PCRDT de manière à pouvoir mieux connaître les agents en termes de caractéristiques mais aussi en termes de collaborations et de partenariats de diverses formes. Les données utilisées pour les estimations présentent l'avantage d'être harmonisées pour l'ensemble des agents quel que soit leur type. De plus, ce sont des données fiables provenant d'une seule source : l'ANRT. Le fait que ces données soient toutes issues de la même base de données est aussi un inconvénient. En effet, pour certaines variables, notamment celles portant sur les moyens en R&D, il aurait été préférable de

bénéficier de données extérieures et plus diversifiées sur les moyens que possèdent chaque structure pour mener à bien leur activité de R&D (dépenses en R&D, nombre de chercheurs, ...). La difficulté est de trouver des données harmonisées pour chaque type d'agents. De la même manière, les données sur la proximité géographique pourraient être améliorées à deux niveaux. Premièrement, en calculant les distances exactes entre les deux agents et non plus entre les grandes villes proches des deux agents. Deuxièmement, il faudrait utiliser une autre mesure du temps de trajet que le temps de parcours par la route de manière à prendre en compte les différents moyens de transport. Enfin, les données sur les réseaux pourraient être complétées par des collaborations effectuées dans d'autres réseaux. Ainsi, on introduirait des variables de contrôle supplémentaires dans le modèle qui prendraient en compte si les agents ont collaboré ensemble dans d'autres formes de réseaux que les PCRDT. ce type de travail pose deux types de difficultés. La première est de posséder suffisamment de base de données sur les collaborations afin de posséder un nombre suffisant de variables de contrôle. La seconde est qu'il faut faire attention à l'endogénéité que ces variables peuvent introduire dans le modèle.

Conclusion générale

A travers l'étude des collaborations au sein des PCRDT, nous avons montré que l'analyse des réseaux sociaux permettait de mieux appréhender ces réseaux complexes que sont les réseaux de collaborations en R&D.

Au sein de la thèse, nous avons étudié les propriétés (générales et individuelles) des réseaux de collaborations au sein des PCRDT, grâce aux outils comme les indicateurs de centralité. Par rapport à la littérature existante, nous avons cherché à mettre en place de nouveaux indicateurs qui permettent de contourner certaines difficultés existantes. De plus, si la plupart des études sur PCRDT sont réalisées dans le cadre des réseaux binaires entre agents, dans la thèse, nous avons cherché à montrer qu'il était important de prendre en compte la répétition des collaborations entre les agents. Enfin, nous avons complété notre analyse en étudiant de réseaux plus agrégés que sont les réseaux de régions qui prennent forme au sein des programmes cadres. Grâce à l'analyse des réseaux sociaux, nous avons montré que les réseaux de collaborations individuels dans les PCRDT possédaient des propriétés favorables à l'innovation. Les programmes cadres tendent donc à devenir cet EER souhaité par l'Union Européenne.

Nous sommes intéressés aux déterminants des choix de collaborations au sein de ces mêmes réseaux, grâce à l'économétrie de réseaux. Nous avons cherché à mettre en exergue quels étaient les principaux déterminants des choix de collaborations des agents au sein du réseau. Nous avons étudié les choix de collaborations pour ensemble d'agents qui possédaient une caractéristique particulière qui était d'avoir participé aux 5^{ème} PCRDT et 6^{ème} PCRDT. Nous avons montré que les choix de collaboration dans les PCRDT étaient influencés par les déterminants traditionnels. En plus de ces déterminants, nous avons mis en avant le rôle important joué par la proximité sociale. Ces déterminants se sont révélés pertinents pour expliquer l'existence d'une collaboration entre deux agents, l'intensité des collaborations ou encore la formation des triades. En outre, les PCRDT sont caractérisés par une diversité des agents qui se traduit par une hétérogénéité des comportements. S'il existe des dispositifs pour favoriser certains types d'agents (PME ou agents

des nouveaux États membres), il pourrait être intéressant de mettre en place des dispositifs incitatifs différents en fonction du type des agents.

Nous avons montré que l'ensemble des études menées au sein de la thèse pouvait être utile dans un objectif d'évaluation des politiques publiques d'innovation, donc dans notre cas d'évaluation des PCRDT. Notons que l'analyse des réseaux sociaux ne permet cependant pas d'évaluer de manière globale ces politiques. Nous avons notamment souligné que l'analyse des réseaux sociaux ne permettait pas de vérifier l'impact des collaborations des PCRDT sur l'innovation en Europe. L'analyse des réseaux sociaux se positionne comme un outil complémentaire susceptible d'apporter des informations pertinentes sur la manière dont sont organisés les réseaux de collaborations et sur la différence entre la réalisation effective de la politique des PCRDT matérialisée par les réseaux de collaborations et les objectifs initiaux assignés à la politique des PCRDT. En construisant des réseaux de régions à partir des réseaux d'agents, nous avons pu mettre en lumière les régions motrices et les axes de collaborations principaux au sein des PCRDT. Nous avons ainsi mis en avant que les PCRDT respectaient la géographie de l'innovation européenne actuelle et que si les régions des nouveaux États membres participaient de plus en plus aux PCRDT, leur participation restait relativement faible et leurs agents avaient des difficultés à faire partie des axes de collaboration privilégiés.

L'analyse des réseaux sociaux présente l'avantage de pouvoir s'adapter au cadre étudié. En effet, l'ensemble des analyses menées dans cette thèse peuvent être utilisées pour étudier d'autres formes de réseaux de collaborations que ce soit d'autres collaborations en R&D tel que les pôles de compétitivité en France ou des réseaux de collaborations différents tel que les réseaux de co-auteurs, les co-brevets,...

Afin d'améliorer et de poursuivre les travaux de la thèse, nous proposons plusieurs pistes de travail. Nous distinguons trois axes de travail principaux qui peuvent être découpées en plusieurs pistes de travail : un axe méthodologique qui vise à améliorer la construction et l'analyse des réseaux et deux axes empiriques. Le premier consiste en un approfondissement des travaux menés dans le cadre de la thèse grâce au recours à de nouvelles bases de données. Le second est une nouvelle piste de travail complémentaire des études menées où l'on cherche à déterminer l'impact des PCRDT sur l'innovation.

Les premières pistes d'approfondissement de la thèse se situent à un niveau méthodologique.

La première consiste à améliorer la construction des réseaux étudiés dans les PCRDT. En effet, dans l'absolu, il faudrait donc pouvoir déterminer les re-

lations précises au sein de chaque projet et notamment pouvoir différencier au sein de chaque projet les collaborations étroites des simples co-participations au même projet. Cette amélioration aurait des impacts à plusieurs niveaux. Sur l'étude des propriétés générales des réseaux, nous aurions alors des réseaux de collaborations "réelles" qui permettraient de relâcher l'hypothèse forte que l'ensemble des agents au sein d'un projet collaborent. De cette manière, nous pourrions distinguer au sein du réseau plusieurs niveaux de collaborations. Dans l'étude des choix de collaborations, il faudrait alors avoir recours à des modèles différents pour distinguer collaboration étroite ou co-participation à un projet ou alors utilisé des modèles de choix multinomiaux ordonnés permettant de distinguer plusieurs choix de collaborations possibles. Obtenir une telle information pour chaque projet apparaît difficile. Une première approche consisterait à utiliser les données financières pour distinguer au sein de chaque projet les agents les plus impliqués et supposés alors que les collaborations étroites ne se déroulent qu'entre ces agents. Pour cela, il faudrait que les données financières soient mieux renseignées au sein de la base.

La seconde piste de travail consiste à poursuivre la réflexion menée sur l'aspect méthodologique qui comprend à la fois la production de nouveaux indicateurs et sur la représentation la plus adaptée aux réseaux de collaborations des PCRDT. Dans le cadre de la thèse, la représentation en "one-mode network" est apparue comme la plus pertinente dans la mesure où nous souhaitons, à la suite de l'étude des réseaux menés dans la partie I, étudier les déterminants des choix de collaborations entre agents. Lorsque l'on souhaite étudier uniquement la centralité des agents, il peut apparaître plus pertinent d'utiliser des "two-mode networks" afin d'éviter de faire l'hypothèse forte que l'ensemble des agents d'un même projet collaborent les uns avec les autres. De manière complémentaire, il peut être intéressant de réfléchir aux indicateurs utilisés en fonction de l'objet et des réseaux étudiés pour prendre en compte certaines particularités.

Le second type d'approfondissement consiste en une poursuite des travaux empiriques menées dans la thèse au travers d'utilisation de données supplémentaires ou de bases de données extérieures. Cette poursuite des travaux empiriques pouvant menées à l'utilisation de nouveaux modèles économétriques.

Dans ce cas, une première piste de travail consiste à étendre le champ d'étude au sein des PCRDT en utilisant un plus grand nombre de programmes cadres et/ou un plus grand nombre de domaines technologiques (ou de thématiques) en réalisant une demande auprès du Ministère de la Recherche et de l'Industrie afin de pouvoir récupérer les données nécessaires.

L'extension du nombre de programmes cadres permettrait d'avoir recours à des modèles de panel qui permettrait de prendre en compte de façon plus précise les phénomènes dynamiques et notamment le renouvellement des collaborations entre deux agents dans le temps. La prise en compte d'un plus grand nombre de domaines technologiques permettrait d'observer s'il existe des différences sectorielles. Le croisement des deux amènerait à une meilleure connaissance des PCRDT. De plus, les estimations se feraient en complément de l'étude des réseaux de chaque programme cadre et /ou de chaque domaine technologique qui permettrait de prendre en compte la trajectoire des agents au sein des programmes cadres et/ou le caractère pluri-disciplinaire des agents. De manière complémentaire, il serait intéressant d'étudier l'ensemble des agents et non seulement ceux qui ont participé aux deux PCRDT. Pour l'instant, les estimations n'ont porté que sur les agents ayant participé au 6^{ème} PCRDT. De cette manière, nous pourrions chercher à montrer que les choix de collaborations sont influencés par la composition du binôme. Selon que les deux agents du binôme aient participé au 5^{ème} puis au 6^{ème} PCRDT, qu'un seul des deux agents du binôme ait pris part aux deux programme cadre ou que les deux sont des nouveaux arrivants, on peut supposer que cela influence les décisions de collaborer ensemble.

Dans le cadre de la thèse, nous avons mis en avant un certain nombre de déterminants susceptibles d'influencer les choix de collaboration. Une autre piste consisterait alors à chercher à améliorer la construction des variables rendant compte de certains déterminants, la proximité géographique ou les moyens en R&D par exemple, ou de chercher à estimer de nouveau déterminants comme les différences sectorielles). Pour cela, il est nécessaire de disposer de données extérieures aux PCRDT. Les thématiques et/ou les domaines technologiques des PCRDT sont des transversales et, de fait fournissent, une information certes importante mais incomplète sur les secteurs d'activité des agents. L'introduction de variables et donc de déterminants au sein des estimations construites à partir de bases de données extérieures aux PCRDT aurait deux effets principaux. Le premier serait de disposer de données plus robustes et moins influencées par les particularités des PCRDT. Dans le cas des moyens en R&D, nous disposerions d'une vue globale sur les dépenses en R&D des agents et non plus seulement des moyens mis en œuvre au sein des PCRDT. De cette manière, il serait possible de distinguer de manière plus importante la différence entre les structures par exemple entre PME et les autres entreprises. Le deuxième effet serait de pouvoir recourir à des estimations complémentaires à celles réalisées dans la thèse mais qui n'ont pu être menées faute de données suffisantes comme un modèle en deux étapes de Heckman dans lequel nous chercherions, dans un première temps, à expliquer l'existence de collaborations puis, dans un second temps, à expliquer

l'intensité des collaborations pour les agents ayant décidé de collaborer. Les principales difficultés pour cet axe de travail sont la disponibilité des données et l'harmonisation de ces données au niveau européen.

Le dernier type d'approfondissement consiste en une extension des travaux menés dans la thèse. Plus particulièrement, nous chercherions à montrer l'impact des réseaux de collaborations des PCRDT sur l'innovation et sur les participants aux PCRDT. Dans ce cas, les PCRDT ne seraient plus alors l'objet d'étude mais un des éléments d'analyse susceptibles d'expliquer les performances en termes de recherche d'agents ou de régions. Par exemple, dans le cas d'un modèle économétrique, les collaborations dans les PCRDT ne seraient plus la variable expliquée mais une des variables explicatives.

Une première difficulté consiste bien sûr à trouver des indicateurs permettant une telle étude. Il est aussi possible de renverser le sens de causalité en considérant les collaborations au sein des PCRDT comme une variable explicative pour expliquer par la suite d'autres formes de collaborations qui sont aussi des productions comme les co-publications et les brevets. Dans ce cas, on chercherait à déterminer dans quelle mesure les PCRDT sont favorables à la production scientifique ou technologique. Ceci permettrait notamment de mieux apprécier si les PCRDT atteignent les objectifs qui leur sont fixés puisqu'ils ont aussi pour but essentiel la production d'innovation. Cette analyse pourrait de plus être articulée avec l'analyse des propriétés générales du réseau afin de déterminer par domaine technologique quelles sont les propriétés du réseau qui favorisent la production scientifique. De cette manière, nous pourrions nous intéresser à la mesure de la performance du réseau. Les brevets et les publications scientifiques mêmes s'ils ont imparfaits peuvent être des premières mesures. Il existe un décalage entre le moment où la collaboration se déroule et le moment où celle-ci lieu à une production (brevet, publication, ...). Ce décalage est souvent compris entre un et trois ans ce qui signifie qu'actuellement, nous sommes en mesure de connaître les répercussions des projets s'étant déroulé au début du 5^{ème} PCRDT. Pour ceux ayant commencé en 2005 et 2006, cela semble encore un peu tôt dans la mesure où en plus du décalage il faut prendre en compte la durée du projet. Il serait intéressant ainsi de connaître si les collaborations dans les PCRDT ont donné lieu à des des co-publications ou des co-brevets. Il serait alors possible de construire un modèle qui cherche à expliquer la formation des réseaux de co-publications par exemple en retenant comme déterminants potentiels les collaborations s'étant déroulé dans les PCRDT. Bien sûr, il faudrait être attentif dans l'estimation au biais d'endogénéité. En effet, les collaborations au sein du 5^{ème} PCRDT peuvent elles-mêmes trouver leur source dans des co-publications précédentes. Une autre application possible serait de déter-

miner si, au niveau français, l'appartenance de deux agents à un pôle de compétitivité ou même à un projet au sein du pôle a favorisé ou non une collaboration de ces agents dans le 7^{ème} PCRDT, étant donné qu'un des objectifs des pôles est la mise en relation d'acteurs au service de l'innovation et de la compétitivité.

Les agents lorsqu'ils participent aux PCRDT cherche avant tout à améliorer leurs performances, dans le cas des entreprises leur profit. C'est aussi l'objectif premier des PCRDT qui ont été créés dans l'objectif d'améliorer la compétitivité des firmes européennes. Dans cette optique, nous pourrions étudier l'impact des PCRDT sur les profits des firmes et notamment comparer l'évolution de ces firmes à celles d'autres firmes européennes semblables n'ayant pas participé aux PCRDT ou encore à des firmes japonaises ou américaines du même secteur. De la même manière que nous avons mis en avant des différences de comportement entre agents au sein des PCRDT, il apparaît opportun de ne pas retenir la même mesure de performance pour évaluer l'impact des PCRDT sur la performance des agents. Ainsi, dans le cas des centres de recherche et/ou des agents de l'enseignement supérieur, la notion de profit est plus difficilement utilisable (et mesurable). Pour ce type d'agents, l'impact des PCRDT sur un potentiel scientifique comme les publications par exemple s'avère plus pertinent.

Enfin, il est important de noter que si une telle évaluation peut être menée au niveau des agents, il est possible de la mener à un niveau plus agrégé notamment au niveau des régions et des pays à condition de trouver une mesure pertinente de la performance à ce niveau (l'indicateur de performance d'Eurostat peut être l'une d'entre elles).

Bibliographie

- [1] Abbess C., Jarrett D. et Wright C.C. (1981), “Accidents at blackspots : estimating the effectiveness of remedial treatment, with special reference to the “regression-tomean” effect”, *Traffic Engineering and Control*, **22** (10) : 535-542.
- [2] Adamic L.A. et Huberman B.A. (2002), “Zipf’s law and the Internet”, *Glottometrics*, **3** : 143-150.
- [3] Agresti. A (1990), “Categorical data analysis”, New-York, John Wiley.
- [4] Allain E. et Brenac T. (2001), “Modèles linéaires généralisés appliqués à l’étude des nombres d’accidents sur des sites routiers : le modèle de Poisson et ses extensions”, *Recherche Transports Sécurité*, **72** : 3-18.
- [5] Almeida P. et Kogut B. (1999), “Localization of knowledge and the mobility of engineers in regional networks”, *Management Science*, **45** (7) : 905-918.
- [6] Archambault J.P. (2004), “Les TIC, la Formalisation et le Partage des Savoirs. Vers une Économie de la Connaissance.”, *Medialog*, **49**.
- [7] Arranz N. et Fernandez de Arroyabe J.C. (2008), “The choice of partners in R&D cooperation : An empirical analysis of Spanish firms”, *Technovation*, **28** : 88-100.
- [8] Audretsch D.B. et Feldman M.P. (2004), “Knowledge Spillovers and the Geography of Innovation” dans *Vernon Henderson J. et Thisse J.F. (eds), Handbook of Regional and Urban Economics, Elsevier (Amsterdam)*, **4** : 2713-2739.
- [9] Autant-Bernard C., Billand P., Frachisse D. et Massard N. (2007), “Social Distance versus Spatial Distance in R&D Cooperation : Empirical Evidence from European Collaborations Choices in Micro and Nanotechnologies”, *Papers in Regional Science*, **86** (3) : 495-519.
- [10] Bala V. et Goyal S. (2000), “A non-cooperative model of network formation”, *Econometrica*, **(68)** : 1181-1229.
- [11] Barabasi A.-L. et Albert R. (1999), “Emergence of Scaling in Random Network”, *Science*, **74** : 1403-1417.

-
- [12] Barabasi A.-L. et Albert R. (2000), "Statistical mechanics of complex networks", *Review of Modern Physics*, **74** (1) : 47-97.
- [13] Baranes E. et Tropeano J.P. (2003), "Why are Technological Spillovers Spatially Bounded", *Regional Science and Urban Economics*, **33** : 445-466.
- [14] Berge C. (1989), "Hypergraphs : Combinatorics of Finite Sets", Amsterdam : North-Holland Mathematical Library.
- [15] Bertacchini Y. et Dou H. (2001), "The Territorial Competitive Intelligence : A network Concept", Actes du Colloque VSST, Barcelone 15-19 octobre.
- [16] Billand P. et Bravard C. (2004), "Noncooperative Networks in Oligopolies", *International Journal of Industrial Organisation*, **22** : 693-735.
- [17] Billand P., Frachisse D. et Massard N. (2008), "The Sixth Framework Program as an Affiliation Network : Representation and Analysis", Working Papers.
- [18] Bloch F. et Jackson M.O. (2006), "Definitions of equilibrium in networks formation games", *International Journal of Game Theory*, **22** : 693-735.
- [19] Bonacich P. (1972), "Factoring and Weighting Approaches to Status Scores and Clique Identification", *Journal of Mathematical Sociology*, **2** : 113-120.
- [20] Bonacich P. (1978), "Using Boolean Algebra to Analyse Overlapping Memberships", *Sociological Methodology*, **9** : 101-115.
- [21] Bonacich P. (1987), "Power and Centrality : A Family of Measures", *American Journal of Sociology*, **92** : 1170-82.
- [22] Bonacich P. (1991), "Simultaneous Group and Individual centralities", *Social Networks*, **13** : 155-168.
- [23] Borgatti S.P. et Everett M.G. (1997), "Network of Analysis of 2-Mode Data", *Social Networks*, **19** : 243-269.
- [24] Borgatti S.P. et Everett M.G. (1999), "Models of core/periphery structures", *Social Networks*, **21** : 375-395.
- [25] Borgatti S.P. et Everett M.G. (2000), "Peripheries of cohesive subsets", *Social Networks*, **21** : 397-407.
- [26] Borgatti S.P., Everett M.G. et Freeman L.C. (2002), *Ucinet for Windows : Software for SocialNetwork Analysis*, Harvard, M.A. : Analytic Technologies.
- [27] Borgatti S.P. (2006), "Identifying set of key players in a social network", *Computational & Mathematical Organization Theory*, **12** (1) : 21-34.

- [28] Boyd J.P., Fitzgerald W.J. et Beck R.J. (2006), "Computing core/periphery structures and permutation tests for social relation data", *Social Networks*, **28** (2) : 165-178.
- [29] Breschi S. et Catalini C. (2010), "Tracing the linkages between science and technology : An exploratory analysis of the research networks among scientists and inventors", *Research Policy*, **39** : 14-26.
- [30] Breschi S. et Cusmano L. (2004), "Unveiling the structure of an european research aera : Emergence of oligarchic networks under E.U. Framework Programmes", *International Journal of Technology Management*, **27** : 747-782.
- [31] Breschi S. et Lissoni F. (2009), "Mobility of inventors and networks of collaboration : An anatomy of localised knowledge flows.", *Journal of Economic Geography*, **9**(4) : 439-468.
- [32] Burt R.S. (1983), "Cohesion versus structural equivalence as a basis for network subgroups", *Applied network analysis : A methodological introduction*, Beverly Hills : Sage.
- [33] Burt R.S. (1992), "Structural holes : The structure of competition", New York : Academic Press.
- [34] Burt R.S. (1994), "The social structure of competition", *Havard University Press*, Cambridge, M.A.
- [35] Cameron C. et Trivedi P. (1990), "Regression Based Tests for Overdispersion in the Poisson Model", *Journal of Econometrics*, **46** : 347-364.
- [36] Cameron C. et Trivedi P. (1998), "Regression Analysis of Count Data", Cambridge, University Press.
- [37] Carayol N. et Roux P. (2008), "The strategic formation of inter-individual collaboration networks. Evidence from co-invention patterns", *Annales d'Économie et de Statistique*, **87-88** : 275-301.
- [38] Carayol N. et Roux P. (2009), "Knowledge flows and the geography of networks : A strategic model of small world formation", *Journal of Economic Behavior & Organization*, **71**(2) : 414-427.
- [39] Cassier M. et Foray D. (1999), "La régulation de la propriété intellectuelle dans les consortiums de recherche : les types de solutions élaborées par les chercheurs.", *Economie appliquée*, Tome LII.
- [40] Cassiman B. et Veugelers R. (2002), "R&D Cooperation and Spillovers : Some Empirical Evidence from Belgium", *American Economic Review*, **92** (4) : 1169-1184.
- [41] Chow G. (1960), "Tests of Equality Between Sets of Coefficients in Two Linear Regressions", *Econometrica*, **28** (3) : 591-605.

-
- [42] Cloudt M., Hagedoorn J. et Roijakkers N. (2006), "Trends and patterns in inter-firm R&D networks in the global computer industry : a historical analysis of major developments during the period 1970-1999.", *Business History Review*, **80** : 725-746..
- [43] Cohen R. et Havlin S.(2003), "Scale free networks are ultrasmall", *Physical Review Letters*, **91** (5).
- [44] Cohen W. et Levinthal D.A. (1990), "Absorptive capacity : a new perspective on learning and innovation", *Administrative Science Quarterly*, **35** : 128-152.
- [45] Coleman J. (1966), "Medical innovation : A diffusion Study", Second Edition, Bobbs-merrill, NY.
- [46] Colletis G. et Winterhalter F. (1991), "Coopération Internes et Externes des Nouveaux Districts Technologiques", Colloque de l'Association des sciences régionales de langue française, Montréal.
- [47] Commission Européenne (2002), "The Sixth Framework Program in Brief".
- [48] Cornwell B. (2005), "A Complement Derived Centrality Index for Disconnected Graphs", *Connections*, **26** (2) : 70-81.
- [49] Cowan R. et Jonard N. (2004), "Network structure and diffusion of knowledge", *Journal of Economic Dynamics and Control*, **28** (8) : 1557-1575.
- [50] D'Aspremont C. et Jacquemin A. (1988), "Cooperative and non-cooperative R&D in duopoly with spillovers", *American Economic Review*, **78** : 1133-1137.
- [51] Deroian F., M'Chirgui Z. et Milleli C. (2007), "Evidences on inter-firm R&D partnerships in three high-tech industries", GREQAM Working Paper.
- [52] Doreian P. (1988), "Equivalence in a social network", *The Journal of Mathematical Sociology*, **13** : 243-282.
- [53] Dosi et Teece (1997), " ", *Nature*, **393** : 440-442.
- [54] Duysters G. et Hagedoorn J. (2002), "External Appropriation of Innovation Capabilities : The Choice Between Strategic Partnering and Mergers and Acquisitions.", *Journal of management studies*, **39** (2) : 167-188.
- [55] Easley D. et Kleinberg J. (2010), "Networks, Crowds and Markets : Reasoning about a Highly Connecting World", *Cambridge University Press*.
- [56] Englmann F. et Walz U. (1995), "Industrial Centers and Regional Growth in the Presence of Local Inputs", *Journal of Regional Science*, **35** (1) : 3-27.

-
- [57] Erdős P. et Rényi A. (1959), "On Random Graphs. I.", *Publicationes Mathematicae*, **6** : 290-297.
- [58] Erdős P. et Rényi A. (1960), "On the Evolution of Random Graphs", *Mathematical Institute of the Hungarian Academy of Sciences*, **5** : 17-61.
- [59] Eurostat (2007), "Science, technology and innovation in Europe".
- [60] Fafchamps M., van der Leij M.J. et Goyal S. (2010), "Matching and network effects", *Journal of the European Economic Association*, à paraître.
- [61] Faust K. (1997), "Centrality in Affiliation Networks", *Social Networks*, **19** : 157-191.
- [62] Foray D. (2000), "L'économie de la connaissance", *La découverte, Collection Repères*.
- [63] Frachisse D. (2007), "L'analyse réseau comme outil d'évaluation de la politique d'innovation : Application au domaine des nanotechnologies dans le cadre des PCRD", Working Papers.
- [64] Freeman L.C. (1979), "Centrality in Social Networks : Conceptual Clarification", *Social Networks*, **1** : 215-239.
- [65] Godfrey L.G. (1978), "Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables", *Econometrica*, **46** : 1293-1302.
- [66] Goerzen A. (2007), "Alliance Networks and Firm Performance : The Impact of Repeated Partnership", *Strategy Management Journal*, **28** : 487-509.
- [67] Gomes-Casseres B., Hagedoorn J. et Jaffe A. (2006), "Do alliances promote knowledge flows ?", *Journal of Financial Economics*, **80 (1)** : 5-33.
- [68] Goyal S. et Joshi S. (2003), "Networks of Collaboration in Oligopoly", *Games and Economic Behavior*, **43 (1)** : 57-85.
- [69] Goyal S. et Joshi S. (2006), "Unequal connections", *International Journal of Game Theory*, **34** : 319-349.
- [70] Goyal S., van der Leij M. et Moraga-Gonzalez J.L. (2006), "Economics : An Emerging Small World", *Journal of political economy*, **114 (2)** : 403-412.
- [71] Graitson D. (2000), "Les Grappes Industrielles : Concept et Méthodologie", Rapport présenté lors de la conférence wallonne de l'innovation.
- [72] Granovetter M. (1973), "The Strength of Weak Ties", *American Journal of Sociology*, **78 (6)** : 1360-1380.
- [73] Granovetter M. (1983), "The Strength of Weak Ties : A Network Theory Revisited", *Sociological Theory*, **1** : 201-233.

-
- [74] Greene W.H. (1981), "Sample Selection Bias as a Specification Error : Comment", *Econometrica*, **49** (3) : 745-758.
- [75] Greene W.H. (2005), "Econometric Analysis", *Pearson Education*, 5^{ème} Edition.
- [76] Gulati R. (1993), "The dynamics of alliance formation", Unpublished doctoral dissertation, Harvard University.
- [77] Gulati R. (1995), "Social structure and alliance formation patterns : a longitudinal analysis", *Administrative Science Quarterly*, **40** : 619-652.
- [78] Gulati R. (1998), "Alliances and networks", *Strategic Management Journal*, **19** : 293-317.
- [79] Hagedoorn J., Link A.N. et Vonortas N.S. (2000), "Research partnerships", *Research Policy*, **29** : 567-586.
- [80] Hagedoorn J. (2002), "Inter-firm R&D partnerships : an overview of major trends and patterns since 1960", *Research Policy*, **31** : 477-492.
- [81] Hagerstrand T. (1969), "Innovation diffusion as a spatial process", University of Chicago Press, Chicago.
- [82] Heckman J. (1979), "Sample Selection Bias as a Specification Error", *Econometrica*, **47** (1) : 153-161.
- [83] Hojman D. et Szeidl A. (2005), "Core and Periphery in Endogenous Networks", *KSG Faculty Research Working Paper Series*, RWP06-022, Harvard University.
- [84] Holland et Leinhardt (1971), "Transitivity in structural models of small group", *Comparative Group Studies*, **2** : 107-124.
- [85] Inno Metrics (2006), "European Innovation Scoreboard 2006".
- [86] Jackson M.O. et Wolinsky A. (1996), "A Strategic Model of Social And Economics Networks", *Journal of Economic Theory*, **71** :44-74.
- [87] Jaffe A.B. (1989), "Real Effects of Academic Research", *The American Economic Review*, **79** : 957-990.
- [88] Jaffe A.B., Trajtenberg M. et Henderson R. (1993), "Geographic localization of knowledge spillovers as Evidenced by Patent Citations", *The Quarterly Journal of Economics*, **108** (3) : 577-598.
- [89] Jarvis D.A. (2006), "A methodology for analyzing complex military command and control (C2) networks", Working papers.
- [90] Kwan M.-P., "Space-Time and Integral Measures of Individual Accessibility. A Comparative Analysis Using a Point-Based Framework, *Geographical Analysis*, **30** (3) (1998) 191-216.

- [91] Latora V. et Marchiori M. (2001), "Efficient behavior of small-world networks", *Physical review letters*, **87** : 1-11.
- [92] Latora V. et Marchiori M. (2003), "Economic small-world behavior in weighted network", *The European Physical Journal B*, **32** : 249-263.
- [93] Latora V. et Marchiori M. (2007), "A measure centrality based on the network efficiency", *New Journal of Physics*, **9** : 1-11.
- [94] Leblond S. (2003), "Guide d'économétrie appliquée pour Matlab".
- [95] Lorrain F. et Harrison C. White. (1971), "The structural equivalence of individuals in social networks", *Journal of Mathematical Sociology*, **1** : 49-80.
- [96] Mac Fadden D. (1974), "The measurement of Urban Travel Demand", *Journal of Public Economics*, **3** : 303-328.
- [97] Malerba F., Vonortas N., Breschi S. et Cassi L. (2006), "Evaluation of progress towards a European Research Area for Information Society Technologies", CESPRI Ũ Bocconi University.
- [98] Manski C. et Lerman S. (1977), "The Estimation of Choice Probabilities from Choice Based Samples", *Econometrica*, **45** : 1977-1988.
- [99] Martin P. et Ottaviano G. (1999), "Growing Location : Industry Location in a Model of Endogenous Growth", *European Economic Review*, **43 (2)** : 281-302.
- [100] Massard N., Autant-Bernard C. et Riou S. (2003), "Polarisation des Activités de Recherche et Développement et Diffusion des Connaissances", *Convention d'étude avec le Commissariat Général au Plan*.
- [101] Mc Cullagh P. et Nelder J.A. (1989), "Generalized linear models", 2ème édition, *Monographs on statistics and applied probability*, Chapman et Hall.
- [102] Moreno J.L. et Jennings H.H. (1938), "Statistics of Social configuration", *Sociometry*, **1** :342-374.
- [103] Moreno J.L. et Jennings H.H. (1945), "Sociometric measurement of Social configurations, based on the deviation from chance", *Sociometry Monograph*, Num. 3, Beacon House, NY.
- [104] Morgan K. (2004), "The exaggerated death of geography : Learning, proximity and territorial innovation systems", *Journal of Economic Geography* : **4 (1)** : 3-21.
- [105] Newman Mark E. J., Strogatz S. et Watts D. (2001), "Random Graphs with Arbitrary Degree Distributions and Their Applications", *Physical Review*, **E 64** : 1-17.

- [106] Ozman M. (2009), "Inter-firm networks and innovation : a survey of literature", *Economic of Innovation and New Technology*, **18** (1) : 39-67.
- [107] Paier M.F. et Scherngell T. (2008), "Determinants of Collaboration in European R&D Networks : Empirical Evidence from a Binary Choice Model Perspective", Working Papers.
- [108] Ponds R., van Oort F. et Frenken K. (2007), "The Geographical and Institutionnal Proximity of Scientific Collaboration Networks", *Papers in Regional Sciences*, **86** : 423-443.
- [109] Porter M. (1990), "The competitive advantage of nations", New York : Free Press.
- [110] Powell W.W., Koput K.W., Smith-Doerr L. et Owen-Smith J. (1999), "Network Position and Firm Performance : Organizational Returns to Collaboration in the Biotechnology Industry", *Research in the Sociology of Organizations*, **16** : 129-159.
- [111] Raymond W., Mohnen P., Palm F. et van der Loeff S.S. (2010), "Persistence of Innovation in Dutch Manufacturing : Is It Spurious?", *The Review of Economics and Statistics*, **92** (3) : 495-504.
- [112] Réka A. et Barabasi A.-L. (2002), "Statistical mechanics of complex networks", *Reviews of Modern Physics*, **74** : 47-97.
- [113] Roediger-Schluga T. et Barber M.J. (2008), "R&D collaborations networks in the European Framework Program : Data processing, network construction and selected results", *International Journal of Foresight and Innovation Policy*, **4** (3/4) : 321-347.
- [114] Roijakkers N. et Hagedoorn J. (2006), "Inter-firm R&D partnering in pharmaceutical biotechnology since 1975 : Trends, patterns, and networks", *Research Policy*, **35** : 431-446.
- [115] Romer P. (1993), "The economics of new ideas and new goods", Proceedings of the World Bank Annual Conference on Development Economics 1992, World Bank, Washington DC.
- [116] Rossi F. (2005), "Innovation policy in the European Union : instruments and objectives", Working Paper.
- [117] Sailer L. D. (1978), "Structural equivalence : Meaning and definition, computation and application", *Social Networks*, **I** : 73-90.
- [118] Sandström A. et Carlsson L. (2008), "The performance of Policy networks : the Relation between Network Structure and Network Performance", *The Policy Studies Journal*, **36** (4) : 497-524.
- [119] Scherngell T. et Barber M. (2009), "Spatial Interaction Modelling of Cross-region R&D Collaborations. Empirical Evidence from the 5th EU Framework Programmes", *Papers in Regional Science*, **88** (3) : 531-546.

- [120] Singh J.H. (2005), "Collaborative networks as determinants of knowledge diffusion patterns", *Management Science*, **51(5)** : 756-770.
- [121] Spearman C. (1904), "The proof and measurement of association between two things", *American journal of Psychology*, **15** : 72-101.
- [122] Suire R., Vicente J. et Dalla pria Y.(2006), "The Ambivalent Stability Properties of Clusters : Why some Clusters Succed Whereas Others Decline ?", CEREM Working Paper.
- [123] Tholoniati A. (2008), "Intelligence économique territoriale et pôles de compétitivité : contribution méthodologique pour l'aide à la décision publique", Thèse soutenue en décembre 2008.
- [124] Union Européenne (Site Europa, 2002), "Sixième programme cadre (2002-2006)" (Synthèse de la législation).
- [125] Union Européenne (Cordis, 2006), "The importance of research in SMEs".
- [126] Veall M. et Zimmermann K. (1992), "Pseudo- R^2 in the Ordinal Probit Model", *Journal of mathematical sociology*, **16** : 333-342.
- [127] Verspagen B. et Duysters G. (2004), *Technovation*, **24** : 563-571.
- [128] Wagner C.S., Thomson R., Tesch T. et Perez R. (2004), "Evaluation of Networks of Collaboration Between Participants in IST Research and their Evolution to Collaborations in the European Research Area (ERA)", Rapport intermédiaire.
- [129] Wasserman S. et Faust K. (2007), "Social Network Analysis. Methods and Applications.", *Cambridge, University Press*, 1^{ère} édition en 1994.
- [130] Watts D. (1999), "Small Worlds : the Dynamics of Networks between Order and Randomness", Princeton University Press, Princeton, New Jersey.
- [131] Watts et Strogatz (1998), "Collective dynamics of "small-world" networks", *Nature*, **393** : 440-442.
- [132] Zimmermann C. (2007), "Academic Rankings with RePEc", University of Connecticut, Department of Economics Working Paper Series.
- [133] Zucker L.G., Darby M.R. et Armstrong J. (1994), "Intellectual capital and the firm : the technology of geographically localized knowledge spillovers", NBER, Working Paper num. 4946, Cambridge.

Abstract :

The Framework Programs (FPs) are the European Union's (EU's) main tools for the promotion of research and innovation within the European territory. They both improve the competitiveness of European firms and promote the emergence of a European Research Area (ERA). Complex networks arise from the collaborations formed between the actors of research and innovation, through projects funded by the FPs.

The objective of this thesis is first of all to show how network analysis provides an interesting perspective for a better understanding of FPs and then to assess the impact of innovation policies within the EU. This analysis will then enable an examination of the structural properties of collaborations formed under the FPs as well as a report on the determinants of collaborations within these programs. The work in this thesis focuses upon the Information Society Technology (IST) of the 6th FP. Two technological fields are used : Telecommunication and Electronics-Microelectronics.

Initially, we study the networks formed by collaborations between two agents from selected technological fields and the aggregated network formed by these two fields of research. We study the general properties of these networks. We aim to verify whether or not these networks have the properties of "small-world networks". This type of network is known in the literature to foster innovation networks. We also observe the characteristics of their evolution over time. Then, we study the agents which compose the networks and particularly we highlight the existence of "hubs". We select the appropriate indicators and, where necessary, we develop new indicators as the strengthened degree for example. We conduct the analysis at both macroeconomic and microeconomic level. Finally, we construct networks which are aggregated at the regional level in order to study regional collaborations. Thus, it is possible to observe whether the FPs permit harmonious development within the EU and involve the construction of an ERA.

In a second part, we study the determinants which may induce both agents to collaborate. We are particularly interested in agents who have experience in FPs. We study the role played by the traditional determinants as the means for R&D, geographical proximity, ... In addition, we emphasize the role played by previous networks in collaboration choices. We study both the existence of collaboration between two agents and the intensity of these collaborations. Finally, in order to extend the analysis of the FPs, we study the formation of triads and particularly we show the existence of the phenomenon of "triadic closure".

Department : Economics (Department 05).

Keywords : Social Networks, Innovation, Collaboration, Spillovers, Framework Programs.

Résumé :

Les Programmes Cadres de Recherche et développement technologique (PCRDT) sont les principaux instruments dont dispose l'Union Européenne (UE) pour, d'une part favoriser la recherche et l'innovation sur le territoire européen et améliorer la compétitivité des firmes européennes, d'autre part faire émerger un Espace Européen de la Recherche (EER). Les collaborations formées entre les acteurs de la recherche et l'innovation, grâce aux projets financés par les PCRDT, forment des réseaux complexes.

L'objectif de la thèse est de montrer de quelle manière l'analyse réseau constitue une perspective intéressante pour mieux appréhender les PCRDT et évaluer l'impact des politiques d'innovations de l'UE. Cette analyse permet à la fois d'examiner les propriétés de la structure des collaborations formées dans le cadre des PCRDT et de rendre compte des déterminants des collaborations au sein de ces programmes. Les travaux menés dans la thèse portent sur deux domaines technologiques, issus de la thématique Information Society Technology (IST) du 6^{ème} PCRDT et qui sont des domaines clés de l'innovation dans l'UE : les Télécommunications et l'Electronique-Microélectronique.

Dans un premier temps, nous étudions les réseaux formés par les collaborations entre agents dans les deux domaines technologiques retenus ainsi que dans le réseau agrégé de ces deux domaines. Nous nous intéressons aux propriétés générales de ces réseaux et cherchons à vérifier s'ils possèdent les propriétés des "small-world networks" reconnus dans la littérature comme des réseaux favorisant l'innovation. De plus, nous observons si ces réseaux évoluent de manière positive dans le temps. Après avoir étudié les réseaux de manière globale, nous nous intéressons aux agents qui composent le réseau en montrant notamment l'existence de "hubs". Dans le cadre de cette analyse, nous sélectionnons les indicateurs adéquats et si nécessaire mettons en place de nouveaux indicateurs comme le degré renforcé. Nous menons ainsi une analyse à deux niveaux : au niveau macroéconomique et au niveau microéconomique. Enfin, nous construisons des réseaux agrégés au niveau des régions afin de ne pas étudier seulement les réseaux de collaborations au niveau des agents mais aussi au niveau des territoires. De cette manière, il est possible d'observer dans quelle mesure les PCRDT impulsent un développement harmonieux de l'innovation au sein de l'UE.

Dans un second temps, nous nous concentrons sur les déterminants qui peuvent inciter deux agents à collaborer. Pour cela, nous nous intéressons plus particulièrement aux agents qui ont déjà une expérience dans les PCRDT. Parmi ces déterminants, on retrouve ceux de la littérature traditionnelle comme les moyens en R&D, la proximité géographique, ... En outre, nous mettons également en exergue le rôle prépondérant joué par les réseaux de la période précédente dans les choix de collaboration et la constitution des réseaux présents. Nous étudions à la fois l'existence d'une collaboration entre deux agents mais aussi l'intensité de ces collaborations. Enfin, afin d'approfondir l'analyse menée dans le cadre de collaborations bilatérales, nous étudions la formation de triades en cherchant à montrer l'existence du phénomène de "triadic closure" tout en observant le rôle joué par les déterminants cités précédemment.

Section : Sciences-Économiques (section 05).

Mots clés : Réseaux sociaux, Innovation, Collaboration, Externalités.

Abstract :

The Framework Programs (FPs) are the European Union's (EU's) main tools for the promotion of research and innovation within the European territory. They both improve the competitiveness of European firms and promote the emergence of a European Research Area (ERA). Complex networks arise from the collaborations formed between the actors of research and innovation, through projects funded by the FPs.

The objective of this thesis is first of all to show how network analysis provides an interesting perspective for a better understanding of FPs and then to assess the impact of innovation policies within the EU. This analysis will then enable an examination of the structural properties of collaborations formed under the FPs as well as a report on the determinants of collaborations within these programs. The work in this thesis focuses upon the Information Society Technology (IST) of the 6th FP. Two technological fields are used : Telecommunication and Electronics-Microelectronics.

Initially, we study the networks formed by collaborations between two agents from selected technological fields and the aggregated network formed by these two fields of research. We study the general properties of these networks. We aim to verify whether or not these networks have the properties of "small-world networks". This type of network is known in the literature to foster innovation networks. We also observe the characteristics of their evolution over time. Then, we study the agents which compose the networks and particularly we highlight the existence of "hubs". We select the appropriate indicators and, where necessary, we develop new indicators as the strengthened degree for example. We conduct the analysis at both macroeconomic and microeconomic level. Finally, we construct networks which are aggregated at the regional level in order to study regional collaborations. Thus, it is possible to observe whether the FPs permit harmonious development within the EU and involve the construction of an ERA.

In a second part, we study the determinants which may induce both agents to collaborate. We are particularly interested in agents who have experience in FPs. We study the role played by the traditional determinants as the means for R&D, geographical proximity, . . . In addition, we emphasize the role played by previous networks in collaboration choices. We study both the existence of collaboration between two agents and the intensity of these collaborations. Finally, in order to extend the analysis of the FPs, we study the formation of triads and particularly we show the existence of the phenomenon of "triadic closure".

Department : Economics (Department 05).

Keywords : Social Networks, Innovation, Collaboration, Spillovers, Framework Programs.

Résumé :

Les Programmes Cadres de Recherche et développement technologique (PCRDT) sont les principaux instruments dont dispose l'Union Européenne (UE) pour, d'une part favoriser la recherche et l'innovation sur le territoire européen et améliorer la compétitivité des firmes européennes, d'autre part faire émerger un Espace Européen de la Recherche (EER). Les collaborations formées entre les acteurs de la recherche et l'innovation, grâce aux projets financés par les PCRDT, forment des réseaux complexes.

L'objectif de la thèse est de montrer de quelle manière l'analyse réseau constitue une perspective intéressante pour mieux appréhender les PCRDT et évaluer l'impact des politiques d'innovations de l'UE. Cette analyse permet à la fois d'examiner les propriétés de la structure des collaborations formées dans le cadre des PCRDT et de rendre compte des déterminants des collaborations au sein de ces programmes. Les travaux menés dans la thèse portent sur deux domaines technologiques, issus de la thématique Information Society Technology (IST) du 6^{ème} PCRDT et qui sont des domaines clés de l'innovation dans l'UE : les Télécommunications et l'Electronique-Microélectronique.

Dans un premier temps, nous étudions les réseaux formés par les collaborations entre agents dans les deux domaines technologiques retenus ainsi que dans le réseau agrégé de ces deux domaines. Nous nous intéressons aux propriétés générales de ces réseaux et cherchons à vérifier s'ils possèdent les propriétés des "small-world networks" reconnus dans la littérature comme des réseaux favorisant l'innovation. De plus, nous observons si ces réseaux évoluent de manière positive dans le temps. Après avoir étudié les réseaux de manière globale, nous nous intéressons aux agents qui composent le réseau en montrant notamment l'existence de "hubs". Dans le cadre de cette analyse, nous sélectionnons les indicateurs adéquats et si nécessaire mettons en place de nouveaux indicateurs comme le degré renforcé. Nous menons ainsi une analyse à deux niveaux : au niveau macroéconomique et au niveau microéconomique. Enfin, nous construisons des réseaux agrégés au niveau des régions afin de ne pas étudier seulement les réseaux de collaborations au niveau des agents mais aussi au niveau des territoires. De cette manière, il est possible d'observer dans quelle mesure les PCRDT impulsent un développement harmonieux de l'innovation au sein de l'UE.

Dans un second temps, nous nous concentrons sur les déterminants qui peuvent inciter deux agents à collaborer. Pour cela, nous nous intéressons plus particulièrement aux agents qui ont déjà une expérience dans les PCRDT. Parmi ces déterminants, on retrouve ceux de la littérature traditionnelle comme les moyens en R&D, la proximité géographique, ... En outre, nous mettons également en exergue le rôle prépondérant joué par les réseaux de la période précédente dans les choix de collaboration et la constitution des réseaux présents. Nous étudions à la fois l'existence d'une collaboration entre deux agents mais aussi l'intensité de ces collaborations. Enfin, afin d'approfondir l'analyse menée dans le cadre de collaborations bilatérales, nous étudions la formation de triades en cherchant à montrer l'existence du phénomène de "triadic closure" tout en observant le rôle joué par les déterminants cités précédemment.

Section : Sciences-Économiques (section 05).

Mots clés : Réseaux sociaux, Innovation, Collaboration, Externalités.

UNIVERSITE JEAN MONNET DE SAINT-ETIENNE
Groupe d'Analyse et de Théorie Economique Lyon Saint-Etienne

THESE
Pour obtenir le grade de
Docteur en Sciences Economiques
Présentée et soutenue publiquement par

David FRACHISSE

**STRUCTURES ET DETERMINANTS DES COLLABORATIONS
AU SEIN DES PROGRAMMES CADRES DE RECHERCHE ET
DEVELOPPEMENT TECHNOLOGIQUE DE L'UNION
EUROPEENNE.
UNE PERSPECTIVE RESEAU.**

ANNEXES

Directeurs de thèse:

Monsieur Pascal BILLAND, Maître de conférences, Habilité à diriger des recherches à l'Université Jean Monnet de Saint-Étienne.

Madame Nadine MASSARD, Professeur des Universités à l'Université Jean Monnet de Saint-Étienne.

Rapporteurs:

Monsieur Nicolas CARAYOL, Professeur des Universités à l'Université Montesquieu de Bordeaux.

Monsieur Koen FRENKEN, Professeur des Universités à l'Université d'Utrecht.

Suffragants:

Monsieur Pierre MOHNEN, Professeur des Universités à l'Université de Maastricht.

Monsieur Alain QUEVREUX, Chef du Service Europe de l'ANRT.

Sommaire

A.1	Présentation des PCRDT	6
A.2	Les informations contenues dans les bases PCRDT	18
A.3	Recherche des doublons et harmonisation des variables propres aux agents	23
A.4	Robustesse des réseaux des deux domaines techno- logiques retenus à la suppression des agents les plus cen- traux	29
A.5	Identification des “hubs” dans les deux domaines tech- nologiques retenus	31
A.6	Identification des “hubs” dans le réseau agrégé	39
A.7	Robustesse des réseaux du sous-domaine nanotechno- logies à la suppression des agents les plus centraux	43
A.8	Identification des “hubs” dans les réseaux nanotechnolo- gie	45
A.9	Indicateur de participation des pays	49
A.10	Cœur des réseaux des deux domaines technologiques re- tenus par la méthode des seuils	55
A.11	Cœur des réseaux des deux domaines technologiques re- tenus par la méthode cœur/périphérie	59
A.12	Performance des régions au sein des PCRDT	61
A.13	Existence des collaborations : Présentation des hypo- thèses	72
A.14	Existence des collaborations : Estimations	78
A.15	Existence des collaborations : Le rôle des collaborations antérieures	87
A.16	Existence des collaborations : L’hétérogénéité des colla- borations	89
A.17	Intensité des collaborations : Estimations	93
A.18	Formation des triades	98

Première partie

L'analyse réseau, un outil d'aide à la décision publique : Application aux PCRDT

Annexes A.1 : Présentation des PCRDT

1. Les PCRDT

1.1. La naissance des PCRDT

Au début des années 80, l'Union Européenne (UE) affiche un retard important en matière de Recherche et Développement (R&D) sur les Etats-Unis et le Japon. Afin de combler ce retard, le conseil de l'UE décide¹ de mettre en place les Programmes Cadres de Recherche et Développement Technologique (PCRDT). Le premier débute en 1984. Le 7^{ème} PCRDT a été lancé pour la période 2007-2013.

Programme Cadre	Période	Financement total en milliards d'euro	Partie Communauté Européenne ² en milliards d'euro
1 ^{er}	1984-1987	-	-
2 nd	1987-1991	-	-
3 ^{ème}	1991-1994	-	-
4 ^{ème}	1994-1998	13,215	-
5 ^{ème}	1998-2002	14,960	13,7
6 ^{ème}	2002-2006	17,500	14,682
7 ^{ème}	2007-2013	50,000	47,25

Tableau A : Les PCRDT (1989-2013)

Ces programmes ont été conçus afin de renforcer les bases scientifiques et technologiques de l'industrie européenne. Ces programmes visent à soutenir la compétitivité de l'industrie européenne. L'Union Européenne part de l'idée que la recherche scientifique et technologique est un élément essentiel pour le fonctionnement et la compétitivité des pays développés, donc des pays européens. Le Traité d'Amsterdam (signé le 2 octobre 1997) insiste sur les apports de la recherche scientifique et technologique pour l'Union Européenne. Ces vertus de la recherche sont repris par la suite dans la note de synthèse d'Europa³ concernant le 6^{ème} PCRDT : *“Il (le développement scientifique et technologique) contribue notamment à la croissance et l'emploi, à la protection des consommateurs et de l'environnement, au renforcement de la compétitivité ainsi qu'à la résolution des grandes questions de sociétés*

¹Résolution du 25 juillet 1983 paru au Journal Officiel numéro C208 du 4 août 1983.

³Le site portail de l'Union Européenne.

(vieillesse démographique, mondialisation, changement climatique, etc.).”.

L’objectif premier des programmes cadres est de rattraper le retard en matière de R&D entre l’Europe d’un côté et le Japon et les Etats-Unis de l’autre. L’Union Européenne souhaite que ce rattrapage se fasse de manière harmonieuse au sein de l’UE comme l’avait indiqué le Centre de REcherche Scientifique et Technique (CREST) selon l’avis que le CREST avait exprimé à l’époque : *“Il importe de promouvoir un développement scientifique et technique équilibré au sein de la communauté.”*. Cette volonté trouve son aboutissement au sommet de Lisbonne en 2000 qui met clairement en avant le souhait de l’UE de voir naître un Espace Européen de la Recherche (EER). La création de cet espace est le principal objectif du 6^{ème} PCRDT. Les objectifs fixés par l’Union Européenne lors de ce sommet sont même plus ambitieux puisque l’UE s’est fixé à l’horizon 2010 *“de devenir l’économie de la connaissance la plus compétitive et la plus dynamique du monde capable d’une croissance économique durable accompagnée d’une amélioration quantitative et qualitative de l’emploi et d’une plus grande cohésion social”*.

Les PCRDT sont le principal instrument de l’UE pour encourager les activités de recherche en Europe. Cela se traduit notamment par des financements importants et qui ne cessent de croître au fur et à mesure des programmes cadres. Pour le 6^{ème} PCRDT, le montant des financements totaux est de 17,5 milliards d’euros pour la période 2002-2006 soit 4 à 5% des dépenses en R&D totale des pays membres. Le tableau A récapitule les financements attribués aux programmes cadres les plus récents. A travers ce tableau, on observe clairement que l’UE a une réelle volonté de construire un EER et d’attirer dans celui-ci un nombre croissant d’agents et si possible des leaders dans chaque domaine.

Après avoir présenté de manière générale les PCRDT, nous nous intéresserons dans la section suivante aux objectifs fixés à ces programmes cadres.

1.2. Le fonctionnement des PCRDT

L’UE a conscience de l’importance de l’activité de recherche pour les États membres mais elle fait aussi le constat qu’il est de plus en plus difficile de mener des activités de RDT (Recherche et Développement Technologique, on parlera par la suite de R&D) de manière isolée plus particulièrement au niveau d’une structure quelle qu’elle soit. Cela est aussi vrai à l’échelle d’un

pays car ces activités sont⁴ :

- complexes et pluri-disciplinaires ;
- coûteuses aussi bien en argent qu'en termes de temps, de matériels et de ressources humaines ;
- demandent une masse critique importante.

A travers les programmes cadres, l'UE souhaite favoriser la collaboration entre agents (entreprises, centres de recherche, université,...) au sein de l'Union Européenne afin de contourner les principales difficultés inhérentes aux activités de R&D. Les collaborations entre agents forment alors de véritables réseaux de recherche à l'échelle européenne. Par les programmes cadres, l'UE accompagne et finance en partie des projets de R&D en Europe dans des domaines qu'elle a préalablement choisis. A chaque programme cadre, la Commission Européenne, en concertation avec les États membres, définit les domaines technologiques qu'elle souhaite mettre en avant. La proposition de la Commission Européenne est ensuite soumise au Conseil Européen pour validation.

Selon le processus suivant, les programmes cadres sont alors découpés en plusieurs appels d'offres concernant plusieurs domaines technologiques. Les agents⁵ se regroupent alors au sein de projet de R&D qu'ils soumettent à la Commission Européenne pour financement. Ces projets sont alors évalués sur des critères scientifiques et techniques et seuls un certain nombre d'entre eux sont financés. L'évaluation porte sur la pertinence et la cohérence du projet dans son ensemble et non sur une évaluation individuelle des agents qui y participent. Cette évaluation est réalisée par des experts dans chacun des domaines retenus. Les projets de R&D sont financés en partie seulement pour éviter que les activités de R&D dépendent uniquement des financements publics. De plus, le projet doit respecter un certain nombre de critères selon le type de projet (appelé instrument, cf. section 2.4. de l'annexe A.1) que l'on souhaite soumettre à la Commission Européenne.

Les PCRDT en favorisant la collaboration dans le cadre de projets de R&D permettent ainsi le partage des coûts et la mutualisation des connaissances. Les projets financés dans le cadre des PCRDT sont composés de plusieurs agents permettant ainsi de partager les coûts et d'atteindre une masse critique. De plus, les programmes cadres permettent de mettre en place des collaborations en R&D entre des agents différents (firmes et/ou instituts de recherches (universités, centres de recherche,...)) qui permettent de partager

⁴Note de la Commission Européenne.

⁵Un même agent peut participer à un ou plusieurs projets dans chaque programme cadre.

des connaissances et favorisent la création d'externalités de connaissances. Les PCRDT mettent en avant des collaborations d'agents issus de domaines différents qui permettent de mieux appréhender la recherche. Parallèlement, les PCRDT permettent à des acteurs du monde de la recherche (Universités, Petites et Moyennes Entreprises,...) qui ne peuvent pas prendre part à des activités de R&D d'envergure faute de moyens (humains, financiers, logistiques,...) de participer aux PCRDT.

1.3. Les PCRDT : une politique critiquée

Les PCRDT sont le principal instrument de la politique d'innovation pour l'Union Européenne. L'une des raisons est qu'ils constituent l'un des principaux moyens d'obtenir des financements pour la recherche en Europe (cf. tableau). Toutefois, il n'échappe pas à certaines critiques.

La première critique concerne le volet administratif. En effet, le processus de sélection peut être plus ou moins long entraînant donc un retard entre le moment où il a été décidé de financer un domaine technologique et le moment où les recherches dans ce domaine débutent effectivement. Ce retard peut entraîner un décalage tel qu'au moment où la recherche est effectuée et surtout le moment où celle-ci aboutira à des résultats concrets, le besoin dans le domaine en question ne sera peut être plus forcément aussi fort ou il aura déjà été satisfait. De plus, le processus de négociation au sein de chaque projet peut être assez important dans la mesure où chaque projet regroupe des agents issus de pays différents. Cette longueur dans les procédures peut entraîner deux conséquences négatives. La première, c'est qu'une partie du temps consacré à la recherche est passée en négociations et, de ce fait, ralentit d'autant les activités de R&D et diminue donc l'ampleur du travail accompli. La seconde, c'est qu'il peut arriver que le marché prennent le pas sur les PCRDT, c'est-à-dire qu'entre le moment où il est décidé de mettre en avant une technologie et le moment où les agents travaillent dessus, cette dernière peut déjà être sur le marché. Par exemple, cela s'est produit avec les logiciels sociaux⁶.

La deuxième critique vient du manque de lisibilité au sein des grands projets dans lesquels sont impliqués un nombre d'agents très important. Dans ces projets, il est difficile d'établir si les agents collaborent tous réellement

⁶Les logiciels sociaux concernent tous les logiciels permettant à un groupe de communiquer plus facilement (Twitter), la construction et la solidification de groupes sociaux, les jeux en ligne massivement multijoueur,...

ensemble au sein de chaque projet ou si des sous-groupes se forment ou bien encore quel est le le degré de collaboration entre les agents et le degré d'implication d'un agent. C'est un aspect qui est très important notamment lorsque l'on cherche à étudier le réseau formé au sein des PCRDT. Ces projets de grande envergure ont pour principal vocation d'atteindre une masse critique et de contourner le caractère complexe et pluri-disciplinaire de la recherche en incitant de nombreux agents à collaborer et à mettre des moyens et des connaissances en commun pour atteindre un objectif commun. Si les agents ne collaborent pas tous réellement ensemble, il semble alors difficile de dire que ces projets de grande envergure sont la solution pour surmonter les difficultés liées à la recherche. Enfin, on peut se demander si au sein de ces projets, il ne figure pas des passagers clandestins qui sont là essentiellement dans le but d'obtenir des financements ou à des fins d'apprentissage. On peut également se demander si certains agents ne sont pas choisis en fonction des critères imposés par l'UE afin d'augmenter les chances du projet d'être sélectionné sans pour autant réellement participer au projet par la suite.

Les PCRDT sont des programmes d'envergure qui visent à soutenir les activités de recherche en Europe. Les PCRDT visent à construire un EER dans le but de rendre plus compétitive l'Europe dans les domaines de R&D. Pour cela, les PCRDT proposent des financements importants afin d'attirer les agents et notamment les grands groupes ou centres de recherche leaders dans leur domaine. Étant donné l'envergure des PCRDT, il est aussi difficile de mettre en place un contrôle efficace au sein de chaque projet.

Chaque programme cadre possède ses propres caractéristiques notamment en termes d'objectifs (bien que ces derniers soient tous tournés en vue de réaliser au mieux les objectifs globaux) et de priorités thématiques. Nous nous intéressons plus particulièrement au 6^{ème} PCRDT.

2. Le 6^{ème} PCRDT

Chaque PCRDT soutient un certain nombre de domaines technologiques et se déroule dans un contexte particulier. Ces éléments influencent alors les objectifs fixés à chaque PCRDT. La comparaison entre deux PCRDT est rendue difficile par le fait que les domaines technologiques soutenus ne sont pas les mêmes mais aussi par les différences en termes de financement total d'un programme à un autre. Dans le cas de notre étude, nous étudierons le 6^{ème} PCRDT. Dans certains cas, nous serons amenés à utiliser des données portant sur le 5^{ème} PCRDT notamment dans la partie économétrique de la thèse.

Dans notre étude, nous nous intéressons principalement au 6^{ème} PCRDT

car c'est le dernier PCRDT pour lequel nous avons les données complètes. Ce programme cadre est le premier qui fait suite au sommet de Lisbonne en 2000, sommet qui met en avant la volonté de l'UE de construire un Espace Européen de la Recherche EER). De plus, il se déroule dans un contexte d'adhésion de nouveaux Etats à l'Union Européenne.

2.1. Les objectifs du 6^{ème} PCRDT

Chaque programme cadre a ses propres objectifs qui tendent tous à améliorer la compétitivité et le niveau de la recherche en Europe. Les objectifs initiaux des PCRDT et le contexte de ce 6^{ème} PCRDT influent sur le choix des objectifs que souhaitent atteindre l'UE au travers de ce programme cadre. L'objectif principal du 6^{ème} PCRDT est de : *“to contribute to the creation of the European Research Area (ERA) by improving integration and coordination of research in Europe which is so far largely fragmented”* (Source : Europa).

De manière parallèle, le 6^{ème} PCRDT doit : *At the same time research will be targeted at strengthening the competitiveness of the European economy, solving major societal questions and supporting the formulation and implementation of other EU policies.* (Source Europa) d'où les objectifs suivants (Source : Europa) :

- soutenir la coopération en matière de recherche ;
- promouvoir la mobilité et la coordination ;
- mettre la recherche et l'innovation au service d'autres politiques de l'UE.

Chacun de ces objectifs est soutenu par des mesures particulières. Les PCRDT sont découpés en trois blocs (cf. Annexes) :

- Bloc 1 : Focusing and integrating European Research (bloc qui contient notamment les 7 priorités thématiques) ;
- Bloc 2 : Structuring the ERA ;
- Bloc 3 : Strengthening the foundation of the ERA.

Dans le cadre de notre étude, nous nous intéresserons seulement aux projets de recherche soumis dans le cadre des priorités thématiques du 6^{ème} PCRDT (Bloc 1) afin notamment d'étudier la mise en place de l'EER. L'objectif de la thèse est d'étudier le réseau de collaboration en R&D formé par le 6^{ème} PCRDT et d'expliquer les liens entre les agents. C'est pour cela qu'il apparaît plus opportun de ne travailler que sur les collaborations qui se déroulent dans le cadre des priorités thématiques et qui constituent le cœur de ce programme cadre. De plus, le 6^{ème} PCRDT comprend des dispositifs

incitatifs afin de favoriser la participation des PME (15% du budget leur est attribué) d'où l'intérêt de voir la place occupée par ces dernières au sein du réseau. En étudiant seulement les collaborations au sein des priorités thématiques et en ne retenant pas les schémas spécifiques aux PME, nous pourrions ainsi voir la participation des PME aux projets collaboratifs sans que cette étude soit biaisée par les participations de Enfin, le 6^{ème} PCRDT se déroule dans un contexte d'adhésion de certains pays à l'UE, il sera alors intéressant d'observer l'intégration de ces derniers au sein du 6^{ème} PCRDT.

Le 6^{ème} PCRDT a pour objectif final la réalisation d'un EER possédant des priorités favorables à l'innovation et à la recherche, c'est-à-dire : un espace intégré impliquant de nombreux agents leaders dans leur domaine et qui collaborent de manière intensive, renforcer la base du 5^{ème} PCRDT et intégrer de nouveaux agents par rapport au 5^{ème} PCRDT (notamment agents innovants comme les PME (ref)). De plus, le contexte de ce 6^{ème} PCRDT fait que l'étendue géographique de cet EER doit être plus importante avec l'intégration des nouveaux États membres.

Maintenant que nous avons présenté les objectifs du 6^{ème} PCRDT, nous présentons dans la sous-section suivante les priorités thématiques qui concernent ce programme cadre.

2.2. Les priorités thématiques du 6^{ème} PCRDT

Le 6^{ème} PCRDT met en avant 7 priorités thématiques qui sont autant de domaines de pointe. Afin d'améliorer la recherche, l'UE cible un certain nombre de domaines technologiques, le plus souvent porteur et qui demande d'important moyens financiers. De cette manière, elle concentre le budget consacré aux PCRDT sur un faible nombre de thématique ce qui permet aux agents concernés par ces champs d'obtenir des financements conséquents pour leur projet. Les agents parviennent alors plus facilement à mener une activité de R&D grâce à des moyens financiers plus concentrés (même si cela pénalise d'autres domaines technologiques). En restreignant le nombre de domaines technologiques financés, l'UE se garantit ainsi de disposer ainsi de financement plus conséquent et de pouvoir financer une plus grande partie des projets. En finançant peu de domaines technologiques et en proposant d'importantes aides, l'UE attire de nombreux acteurs de la recherche dans ces domaines (notamment les leaders) et, dans le même temps, permet aux projets financés d'atteindre une taille critique. En finançant, un plus grand nombre de domaines, les ressources dont disposent l'UE pour conduire à bien sa politique serait alors divisée entre plus de projets, les financements seraient alors moins conséquents. De plus, les programmes seraient moins attractifs. Cela pourrait avoir pour double conséquence de conduire des projets moins

ambitieux qui ne permettrait pas de réelles avancées et simultanément des projets menées par des acteurs moins importants en termes de leadership.

Dans le cas du 6^{ème} PCRDT, les 7 priorités thématiques sont :

- Life sciences, Genomics and Biotechnology for Health ;
- Information Society Technologies ;
- Nano-technologies and nano-sciences, knowledge-based multifunctional materials, new production processes and devices ;
- Aeronautics and Space ;
- Food Quality and Safety ;
- Sustainable Development, Global Change and Ecosystems ;
- Citizens and Governance in a knowledge-based society.

Chacune de ces priorités thématiques est elle-même divisée en plusieurs sous-thématiques (cf. Annexes). D'un programme à l'autre ces thématiques évoluent, ce qui rend les comparaisons entre les programme cadres délicates. Ci-dessous, les priorités thématiques du 5^{ème} PCRDT :

- Quality of Life and management of living resources (Quality of Life) ;
- User-friendly information society (IST) ;
- Competitive and sustainable growth (GROWTH) ;
- Energy, environment and sustainable development (EESD).

Entre les deux programmes cadres, on observe une certaine continuité dans les thématiques mais il est difficile d'établir des correspondances entre les deux programmes cadres. Certaines priorités thématiques du 5^{ème} PCRDT ne semblent pas trouver d'équivalent dans le 6^{ème} PCRDT (Aeronautics and Space par exemple) ou d'autres semblent correspondre à deux thématiques (Citizens and Governance in a knowledge-based society).

Ces priorités thématiques sont les domaines dans lesquelles l'UE souhaite investir et donc dans lesquels elle souhaite attirer des agents innovants. Dans la sous-section suivante, nous présentons les différents agents qui peuvent prendre part aux programmes cadres.

2.3. Quels sont les acteurs susceptibles de participer à ce 6^{ème} PCRDT ?

Les programmes cadres cherchent à attirer un grand nombre d'acteurs innovants de types différents. Au sein des programmes cadres, on retrouve quatre grandes catégories d'agents :

- Les entreprises (IND) ;

- Les centres de recherche (RES) ;
- Les agents de l’enseignement supérieur (HE) ;
- La catégorie “autres” (OTH).

La première catégorie d’agents concernent les entreprises.

Dans les programmes cadres, il existe une distinction entre les Petites et Moyennes Entreprise (PME) et les grandes entreprises. Dans le 6^{ème} PCRDT sont considérées comme PME les entreprises suivantes : “*As SME we understand every organisation which has less than 250 employees, a turnover of no more than 40 M or a yearly balance sheet total of no more than 27 M, and which is not be owned by a large enterprise for more than 25%.*”⁷.

La première remarque que l’on peut faire est que cette définition qui permet de distinguer deux types d’entreprises peut s’avérer plus difficile à appliquer dans certains pays notamment en France où les PME sont souvent des entreprises de moins de 50 salariés (ref). Il existe une distinction plus fine dans laquelle les entreprises de moins de 10 salariés sont des Très Petites Entreprises (TPE). Toutefois, cette distinction n’apparaît pas dans les différentes bases de données portant sur les PCRDT.

Il est important de noter que la définition d’une PME a évolué entre le 4^{ème} PCRDT et le 5^{ème} PCRDT. Dans le 4^{ème} PCRDT, les PME étaient définies comme des entreprises de moins de 500 salariés. Ce changement a été fait dans le but de s’ajuster au mieux à la réalité des faits. En effet 95% des PME qui participaient aux projets CRAFT (Co-operative research projects, projet de soutien aux PME) comptaient moins de 250 salariés et deux tiers des PME qui ont participé aux 4^{ème} PCRDT comptait moins de 50 salariés. De plus, ce changement permettait de s’aligner avec les autres politiques communautaires européennes qui utilisaient cette nouvelle définition depuis 1996.

La deuxième catégorie rassemble les centres de recherche des universités ou encore les instituts de recherche. On s’perçoit rapidement que le fait que les centres de recherche des universités occupent cette catégorie alors qu’il existe un catégorie agents de l’enseignement supérieur posera par la suite un certain nombre de problème lors du traitement et de l’étude du 6^{ème} PCRDT. Cette dernière catégorie regroupe les universités mais aussi les grandes écoles. La catégorie autres rassemblent des agents au profil varié. On y retrouve les cabinets de consultants en grande partie mais aussi les fédérations, associations,...

Le but est ainsi de regrouper au sein des projets financés par l’UE des

⁷Source : CORDIS

agents qui sont structurés de manière différentes. Ces agents peuvent poursuivre des objectifs différents et réaliser une recherche différente. Faire collaborer au sein de projets de recherche des structures différentes permet de concilier des compétences, des connaissances et des approches différentes. Ces différences doivent permettre d'améliorer la recherche au sein des projets. Au sein des projets financés par l'UE, on retrouve souvent des entreprises et des centres de recherche affiliés à des universités par exemple. De cette manière cohabitent, au sein des projets, des agents réalisant une recherche opérationnelle à but commercial et de l'autre des agents réalisant une recherche de type fondamentale qui vise à atteindre des objectifs scientifiques. En incitant des agents aux profils différents à collaborer ensemble, l'UE cherche à contourner les difficultés liées au fait que la recherche est de plus en plus une activité pluri-disciplinaire demandant d'énormes capacités que peu de structures possèdent en interne. La collaboration d'agents différents peut cependant être difficile dans la mesure où tous les agents n'attendent pas la même chose au final des projets et ont des manières de fonctionner différentes.

De plus, les programmes cadres permettent aussi d'intégrer des agents qui possèdent des compétences en matière de R&D mais qui participent peu à des programmes d'envergure faute de moyens ou de possibilités. L'UE met en place un certain nombre de dispositifs incitatifs afin de faciliter la participation de ces agents. C'est le cas notamment pour les PME, 15% du financement total leur est réservé et de nombreux projets permettent aux PME de s'intégrer facilement (cf. section 2.4 de l'annexe A.1). Les PME apportent à la fois une expertise différente et peuvent aussi apprendre au sein des PCRD.

Afin d'éviter que les projets soumis à la Commission Européenne ne soient des projets entièrement supportés par des agents d'une même nationalité, l'UE a fixé un certain nombre de contraintes notamment sur la nationalité des agents au sein d'un projet. La première d'entre elles est que le projet doit contenir au moins deux ou trois agents de pays membres différents (il existe des différences selon l'instrument choisi, i.e. selon le type de projet). De cette manière, l'UE pousse à la collaboration européenne afin d'avoir un EER intégré dans lequel il existe des collaborations entre pays. Cette contrainte correspond aussi à la volonté de l'UE de voir un "développement scientifique et technique équilibré" au sein de la communauté. De plus, cette contrainte, dans un contexte d'adhésion de nouveaux États membres à l'UE, permet aussi d'impliquer des agents issus des nouveaux États membres. En outre, on retrouve dans les PCRD des agents issus de pays non membres de l'UE même si leur participation est marginale.

Nous avons présenté les différents types d'agents susceptibles de participer

aux PCRDT. Ces agents se regroupent au sein de projets qu'ils soumettent à la Commission Européenne pour financement comme nous l'avons vu précédemment. Ces projets peuvent être de différents types appelés instruments, chacun répondant à des besoins et des objectifs différents. Nous présentons ces instruments dans la sous-section suivante.

2.4. Les instruments du 6^{ème} PCRDT

Il existe un certain nombre d'instruments au sein des PCRDT. Nous présenterons ici ceux que nous retrouvons dans les différents appels d'offre concernant les 5 priorités thématiques du 6^{ème} PCRDT mentionnées préalablement.

Les différents instruments du 6^{ème} PCRDT sont :

- Network of Excellence (NoE) ;
- Integrated Projects (IP) ;
- Specific Targeted REsearch Projects (STREP) ;
- Coordination Action (CA) ;
- Specific Support Action (SSA).

Les deux premières formes de projets sont des instruments qui visent à ce que les groupes d'agents ainsi formés atteignent la masse critique suffisante afin de mener à bien leur projet de recherche. Ces deux instruments correspondent à des projets d'envergure qui rassemblent un nombre important d'agents. Les PME sont encouragées à participer à ces deux instruments.

Les projets NOE sont des projets à vocation internationale. Cet instrument a pour objectif d'obtenir des changements structurels dans la manière dont est menée la recherche en Europe dans le domaine concerné. Cet instrument a pour but d'atteindre l'excellence scientifique et technologique dans un domaine à travers l'intégration durable des capacités des participants au projet afin si possible d'aboutir à un leadership européen dans le domaine. La création de connaissances nouvelles est aussi un des objectifs de cet instrument, les deux objectifs étant complémentaires les uns des autres. Ces projets sont construits autour d'au moins trois agents de trois pays différents provenant de trois États membres ou pays associés dont au moins deux États membres et/ou pays candidats.

Les projets IP ont des objectifs légèrement moins ambitieux. Ils visent à accroître la compétitivité de l'Europe dans les domaines technologiques choisis et à créer de nouvelles connaissances afin de créer de nouveaux produits, services, procédés, ... La condition concernant la taille minimum du projet et la nationalité de ses agents sont les mêmes que pour les projets NOE.

Les projets STREP sont moins “ambitieux” que les projets NOE et IP et de taille plus réduite. Ces projets visent à améliorer la compétitivité de l’Europe et cherchent à cibler les besoins de la société actuelle. Ces projets peuvent prendre soit la forme d’un projet de R&D visant à accroître le champ des connaissances dans un domaine ou bien à mettre en avant la viabilité d’une technologie qui, même si on est actuellement persuadé de son bien fondé, ne peut pas être commercialisée dans l’état actuel des choses. Les conditions imposées au sein de cet instrument sont identiques aux deux instruments précédents.

Les projets CA et SSA possèdent des objectifs différents des précédents instruments. Ils sont de plus faible envergure. Les projets SSA sont destinés à stimuler, encourager et faciliter la participation des PME et des structures de recherche de faible taille ou des structures localisées dans les nouveaux États membres. De plus, ils servent à préparer les programmes cadres suivants. Les projets CA concernent notamment l’organisation de conférences, de meetings, . . . Ils servent à promouvoir le travail en réseau et la coordination au sein des projets de recherche.

Chaque instrument possède donc ses propres objectifs. Les agents en fonction de leur aspiration décident donc quelle forme d’instrument il souhaite utiliser. Clairement, les projets NOE, IP et STREP ont un contenu beaucoup plus axé sur le développement des connaissances et sur l’amélioration de la compétitivité de l’UE. Pour autant, les deux autres instruments ne sont pas à écarter des programmes cadres dans la mesure où ces deux instruments participent eux aussi à la création de l’EER et à sa structuration. Les projets SSA et CA permettent de favoriser la participation des agents et de promouvoir les conditions essentielles de réussite des projets de R&D. Par cet intermédiaire, ils permettent aux programmes cadres de mieux fonctionner et ainsi d’atteindre les objectifs qui leur sont assignés.

Le 6^{ème} PCRDT s’est déroulé après le sommet de Lisbonne en 2000, qui a vu émerger la volonté de l’UE de créer un EER, et dans un contexte d’adhésion de nouveaux États membres. Ce contexte autour du 6^{ème} PCRDT fait de ce programme cadre un élément encore plus essentiel de la politique européenne d’innovation.

Annexes A.2 : Les informations contenues dans les bases PCRDT

La priorité thématique IST est découpée en 5 domaines technologiques. Pour chacun des ces domaines, nous connaissons les projets retenus par la Commission Européenne et les agents qui participent à ce projet. Un agent peut participer à plusieurs projets différents en même temps dans plusieurs domaines technologiques et/ou thématiques différentes.

1. Les informations sur les projets

Chaque projet est identifié par un nom et un acronyme qui renseignent sur les objectifs du projet. Pour chaque projet, nous connaissons l'ensemble des agents qui y participent regroupés dans un consortium. Le projet est rapidement décrit en une ou deux lignes. Nous connaissons l'appel d'offre auquel il a été soumis, c'est-à-dire que nous connaissons le domaine technologique, la thématique et la date à laquelle il a été soumis. De plus, nous connaissons l'instrument qui a été choisi par le groupe d'agents pour soumettre le projet. Pour rappel, les instruments possibles sont les suivants :

Instrument	Acronyme
Network of Excellence	NOE
Integrated Project	IP
Specific targeted Research Projects	STREP
Coordination Action	CA
Specific Support Action	SSA

Tableau B : Les différents instruments du 6^{ème} PCRDT.

Les deux premiers instruments (NOE et IP) sont des projets d'envergure qui visent notamment à atteindre une certaine masse critique (en termes de moyens financiers, d'équipe de recherche, de connaissances,...) afin de pouvoir atteindre des objectifs ambitieux. Les projets de type NOE visent à obtenir des changements radicaux dans la manière de mener la recherche dans la thématique concernée tandis que les projets de type IP visent en premier lieu la création de nouvelles connaissances pour améliorer le niveau de la recherche en Europe dans la priorité thématique choisie. Les projets STREP sont proches des projets de type IP mais ils ont des ambitions moins élevées que ces derniers. Enfin, les projets de type SSA et CA sont des actions de soutien à la recherche. Ils servent à promouvoir la recherche et favoriser la dissémination des connaissances.

Nous disposons des financements demandés pour le projet par le consortium et le coût du projet pour le consortium. Les financements correspondent à la partie des coûts du projet que le consortium souhaitent voir financer par la Commission Européenne. La différence entre le coût du projet et le financement demandé sera donc à la charge des participants. Les projets retenus dans le cadre des PCRDT ne sont financés qu'en partie même s'il existe des exceptions pour lesquels le projet est financé en totalité. Il s'agit ici de montant demandé qui peuvent différer de ceux réellement obtenus.

Enfin, nous connaissons le coordonnateur du projet. Le coordonnateur est un des agents du projet qui a un rôle central dans le fonctionnement et le bon déroulement du projet. Il joue un rôle d'interface entre les agents du consortium et la Commission Européenne. Il a notamment en charge la présentation des rapports d'activité auprès de la Commission Européenne.

2. Les informations concernant les agents

Ces informations peuvent être classées en deux grandes catégories :

- Les informations fixes d'une participation à une autre ;
- Les informations évolutives d'une participation à un autre.

2.1. Les informations fixes

Nous possédons pour chaque agent un certain nombre de caractéristiques. Nous connaissons pour chacun des agents son type, c'est-à-dire le type d'activité de l'agent (type 1) et son statut juridique (type 2). Il existe quatre types d'activité au sein des programmes cadres :

Type d'activités	Acronyme
Agent de l'enseignement supérieur	HE
Centre de recherche	RES
Industrie	IND
Autres (Consultants, Fédérations,...)	OTH

et sept statuts juridiques :

Statuts juridiques	Acronyme
Gouvernement	GOV
Organisation privée à but non lucratif ⁸	PNP
Organisation commerciale privée (incluant les consultants)	PRC
Groupement d'intérêt économique	EEGI
Organismes internationaux	INO
Organisation commerciale publique	PUC
Joint Research Centre	JRC

Le type d'activité et le statut juridique servent à caractériser les agents.

Pour chaque agent, nous possédons un certain nombre d'éléments qui permettent de le localiser. En effet, nous connaissons l'adresse, la ville, le code postale et le pays de chaque agent. Ces informations sont normalement constantes au sens où si un agent participe à plusieurs projets, elles n'évoluent pas. Les agents sont censés déclarer le lieu où la recherche se déroule réellement et non le siège social pour une entreprise par exemple.

Nous possédons peu d'informations sur la taille des agents (en termes de chiffres d'affaire, de nombre de salariés, ...). La seule information en notre possession sur la taille de l'agent est la distinction être les Petites et Moyennes Entreprises (PME) et les non PME. Une structure est considérée comme une PME si les 3 conditions suivantes sont remplies⁹ :

- L'entreprise est composée de 250 salariés maximum. Ce seuil est le même pour les trois derniers programmes cadres (FP5, FP6 et FP7) alors qu'il était de 500 salariés pour le 4^{ème} PCRD ;
- Le chiffre d'affaires de l'entreprise n'excède pas 50 millions d'euros (seuil de 40 millions d'euros en 1996) et si elle a moins de 43 millions d'euros inscrits au bilan (27 millions d'euros en 1996). L'augmentation des seuils tient compte de l'augmentation des prix et de la productivité depuis 1996 ;
- L'entreprise n'est pas détenue à plus de 25% par une entreprise non PME.

Étrangement, cette information est renseignée quel que soit le type d'activité de l'agent. Afin d'être cohérent avec la définition d'une PME et avec la définition de l'UE (qui favorise les PME au sein des PCRD), seules les entreprises peuvent être des PME. Les autres structures ont dû renseigner ce critère sur la base de leur effectif mais il est difficile d'en être certain. De même, il apparaît dommageable que certaines entreprises identifiées comme des filiales de grand groupe se déclarent comme PME. Ceci peut entraîner des

⁹Source : ANRT.

biais par la suite notamment lorsque l'on s'intéresse aux comportements des PME car on ne peut pas comparer la stratégie de collaboration d'une "vraie" PME et celle d'une filiale de grand groupe même si celles-ci comptent moins de 250 salariés.

Même si ces informations sont censées être fixes, il est possible qu'elles évoluent d'une participation à une autre souvent pour des raisons exceptionnelles du type déménagement (il ne faut pas oublier que le 6^{ème} PCRDT se déroulent sur 4 ans). Il est aussi possible que des erreurs dans la base de données induisent des changements pour certains champs de la base.

2.2. Les informations évolutives

Les informations évolutives sont peu nombreuses. Elles concernent avant tout le coût supporté par l'agent au sein du projet et la part du financement demandé qui lui serait attribué. Nous avons vu à la sous-section précédente que nous possédions comme information le coût du projet et le financement demandé par le consortium. Nous possédons en plus la répartition du coût entre les agents et la répartition du financement.

Ces informations sont importantes dans la mesure où nous ne possédons pas d'informations sur les moyens en R&D (budget en R&D, nombre de chercheurs,...) dont disposent chaque agent. Ces informations peuvent donc servir de proxys notamment dans le cadre d'une étude économétrique. De plus, ces informations peuvent permettre de repérer au sein de chaque projet l'engagement des agents. Ainsi, on peut supposer que si un agent supporte un coût élevé relativement aux autres alors c'est qu'il est un des agents du projet qui a le plus d'intérêt au sein de ce projet, surtout si ce coût élevé s'accompagne parallèlement d'un financement faible.

Ces informations peuvent de la même manière permettre d'identifier des comportements du type "passager clandestin" voire d'identifier des agents "alibis". Un agent alibi est un agent qui participe au projet afin que le consortium respecte les règles fixées par la Commission européenne sans pour autant réellement participer aux travaux de recherche. La Commission Européenne impose des règles strictes pour financer un projet (nationalité des agents qui composent le projet par exemple). De plus, si l'on observe les objectifs des PCRDT, on s'aperçoit que la participation des PME et l'intégration des États membres font partie des objectifs des PCRDT même si ce ne sont pas les objectifs premiers de ces programmes. Le consortium peut alors avoir intérêt d'intégrer au projet des agents de ce type pour augmenter ses chances de financement. On ne sait pas dans quelle mesure ces incitations jouent. Toutefois, lorsque l'on observe financements, il est étrange de noter que la

répartition est loin d'être égalitaire.

La dernière information évolutive est le coordonnateur du projet. un agent peut participer à plusieurs projets au sein du programme cadre. Chaque projet comporte un coordonnateur. Un agent peut donc être plusieurs fois coordonnateur comme ne jamais l'être.

L'ensemble de ces informations présentes dans les bases de données sont récapitulées dans le tableau suivant :

Nom du Champ	Type de données	Description
PRIORITE	Alphanumérique	Appel d'offre auquel répond le projet
CALL	Alphanumérique	Appel d'offre auquel répond le projet avec l'information sur l'année
NUM	Alphanumérique	Numéro attribué au projet
ACRO	Texte	Acronyme du projet
INST	Texte	Instrument choisi pour le projet
TITLE	Texte	Titre complet du projet
TOTCOST	Numérique	Coût total du projet supporté par le consortium
REQBUDG	Numérique	Financement total du projet demandé par le consortium
DURATION	Numérique	Durée du projet
PARTNUM	Numérique	Nombre attribué au participant dans le projet
PARTNAME	Texte	Nom du participant
PART_ADRS	Alphanumérique	Adresse du participant
PART_TOWN	Numérique	Ville où est localisé le participant
NUTS3	Alphanumérique	Nuts 3 où est localisé le participant
PART_POST	Alphanumérique	Code postale de la ville où est localisé le participant
COUNTRY	Texte	Pays du participant
TYP1	Texte	Type d'activité du participant
TYP2	Texte	Statut juridique du participant
SME	Texte	PME ou non PME
PARTOTCOST	Numérique	Coût supporté par le participant dans le projet
PARTREQBUDG	Numérique	Financement demandé par l'agent dans le projet
THÉMATIQUE	Texte	Thématique auquel appartient le projet
DT	Texte	Domaine technologique auquel appartient le projet
PROGRAMME	Texte	Programme auquel appartient le projet

Tableau C : Informations disponibles dans les bases de données des PCRDT.

Annexes A.3 : Recherche de doublons et harmonisation des variables propres aux agents

1. Définition d'un agent au sein des bases PCRDT

La phase préalable à la construction d'un réseau de collaborations est la connaissance des agents qui le composent. Quelle est la définition d'un agent ?

Les agents peuvent participer à plusieurs projets différents au sein des PCRDT. Un agent est susceptible de se retrouver plusieurs fois dans la base de données. Si d'une participation à une autre, le nom de l'agent ainsi que ces principales caractéristiques étaient parfaitement identiques, il suffirait alors de réaliser une requête de regroupement sous Access ou sous Excel pour identifier l'ensemble des agents de la base. En réalité, l'utilisation d'une telle requête sur la base de données se révélerait inefficace. En effet, d'une participation à une autre, les caractéristiques de l'agent, plus particulièrement son nom, peuvent être différents alors qu'il s'agit pourtant bien du même agent. Un certain nombre de facteurs peuvent expliquer cela :

- Modification réelle des caractéristiques de l'agent entre deux participations (déménagement, changement de statut,...) ;
- Différentes langues utilisées (anglais, français, allemand,...) ;
- Utilisation de lettres ou de symboles propres à un pays (comme le eszet β en allemand) ;
- Permutation de mots au sein du même champ ;
- Des inversions entre les champs (entre la ville et le pays par exemple) ;
- Ajout de divers compléments dans le nom du partenaire (le statut juridique des entreprises par exemple) ;
- Utilisation d'acronyme comme PSA pour Peugeot Société Anonyme) ;
- Erreurs de remplissage ;
- Des données manquantes.

De plus, ces différents problèmes s'appliquent aussi aux variables qui permettent de localiser l'agent ainsi qu'à celles concernant le type. L'ensemble de ces problèmes rend l'identification des agents difficile.

Pour que deux participations différentes puissent être attribuées à un même agent, il faut que dans les deux cas les caractéristiques de l'agent, du nom au type en passant par l'adresse, soit strictement identique. Dans l'idéal, il faudrait donc que huit caractéristiques correspondent strictement. Étant donné l'ensemble des problèmes que l'on rencontre dans la base, il apparaît plus pertinent de se restreindre à seulement deux caractéristiques. Ainsi, nous considérerons que deux participations renvoient au même agent

si et seulement si le nom de l'agent ainsi que sa ville correspondent dans les deux participations.

Pour la localisation, il aurait été préférable de prendre l'adresse complète mais deux raisons nous poussent à retenir seulement le champ ville. Premièrement, l'adresse complète est composée de plusieurs mots et/ou chiffres et rend la probabilité de permutation des mots plus importante que pour le champ ville entraînant de fait un coût de traitement bien supérieur sans pour autant garantir un taux d'erreur inférieur. Deuxièmement, le code postal est plus souvent manquant par rapport aux autres informations concernant la localisation rendant l'identification plus difficile. De la même manière, nous ne prenons pas en compte le type des agents car celui-ci peut évoluer pendant les 5 années que durent le 6^{ème} PCRDT. Par exemple, une entreprise peut participer à ce programme cadre en 2002 en tant que PME et à nouveau en 2006 en tant que non PME suite à un agrandissement de l'entreprise sur la période. De plus, certaines structures peuvent déclarer plusieurs types d'activité car la nomenclature s'avère peu précise (cf. ci-après).

La solution retenue présente un biais. Nous pouvons considérer que deux participations soient le fait du même agent alors qu'en réalité il s'agit de deux agents situés dans la même ville et appartenant à une même structure mais à deux adresses différentes. Nous pouvons supposer que ce biais est croissant avec la taille des villes et avec le nombre de participations des agents de cette ville. Pour diminuer ce risque d'erreur, nous effectuons une vérification manuelle comme nous le verrons par la suite.

Pour les grandes entreprises et les groupes, nous avons considéré que la maison mère et ses filiales constituent autant d'agents différents contrairement à Hanaki, Nakajima et Ogura (2007). Cette distinction est utile afin d'observer la répartition spatiale des agents qui participent aux PCRDT et afin de pouvoir mesurer les effets de la proximité géographique sur les déterminants des choix de collaboration. La maison mère et les filiales sont distinguées à l'aide des informations portant sur la localisation. Par exemple, le siège social de la société Philips se trouvant à Eindhoven et l'entreprise Philips située à Paris en France forment deux agents. Ce critère est appliqué aux autres catégories d'agents, notamment aux centres de recherche. Ainsi, à l'image de l'exemple précédent, le CNRS de Paris et celui de Toulouse forment deux agents différents.

2. Méthode utilisée pour la recherche de doublons

Afin d'optimiser la recherche de doublons, nous adoptons une procédure en 4 temps.

2.1. Étape 1 : Nettoyage de la base

La première étape consiste à supprimer tous les symboles qui pourraient empêcher une requête de reconnaître les doublons au sein de la base. Pour chaque champ important pour l'identification des doublons, nous créons une colonne identique que nous allons par la suite modifier pour optimiser la recherche de doublons. Ainsi, dans cette étape, nous supprimons tous les symboles qui sont propres à certaines langues comme par exemple le “ß” (eszet) allemand. Ces symboles peuvent gêner l'identification des doublons si le nom est écrit une première fois à l'aide des symboles et une seconde fois sans ce symbole. De la même manière, nous supprimons au sein de la base tous les caractères de ponctuation “,” “;” “ ”... Afin d'illustrer la nécessité de ces modifications, prenons l'exemple de l'université de Versailles qui apparaît deux fois dans la base écrit de la façon suivante :

Nom	Identifiant
university versailles cnrs £ prism laboratory	1
university versailles cnrs prism laboratory	2

L'objectif est d'attribuer à chaque agent un identifiant qu'il aura pour chacune de ces participations. Ci-dessus, nous avons réalisé une requête qui a pour but de rechercher les doublons et de donner le même identifiant si deux entrées sont identiques. Ici, la requête considère qu'il s'agit de deux entrées différentes à cause du symbole “£”. Si on supprime ce symbole en le remplaçant par rien on obtient le résultat suivant :

Nom	Identifiant
university versailles cnrs prism laboratory	1
university versailles cnrs prism laboratory	2

La requête n'attribue toujours pas le même identifiant à ces deux entrées qui sont pourtant identiques. Ceci est dû à la présence dans la première ligne d'un double espace entre “cnrs” et “prism”. En supprimant les symboles de ponctuation ainsi que les espaces on aboutit au résultat suivant.

Nom	Identifiant
universityversaillescnsrprismlaboratory	1
universityversaillescnsrprismlaboratory	1

La requête attribue enfin le même identifiant à ces deux entrées.

2.2 Étape 2 : Requête pour identifier de doublons

La seconde étape, après le nettoyage des caractères spécifiques et des ponctuations, consiste à appliquer sur les deux champs que sont le nom et la ville une requête qui vise à rechercher les doublons et à attribuer aux agents des identifiants. Ces requêtes sont réalisées à l'aide des logiciels Access et Excel. L'ensemble des symboles à éliminer et la requête sont répertoriées dans l'article de Laurent (2006). La base que l'on obtient en sortie est bien sûr imparfaite et ne peut pas servir en l'état à l'étude des PCRDT. En effet, un certain nombre de problèmes persistent encore comme les inversions de mots, les fautes de frappe et d'orthographe lors de la saisie, l'utilisation de langues différentes,...

2.3. Étape 3 : Utilisation d'algorithme basé sur les distances entre mots

La troisième étape a pour objectif de résoudre en partie les problèmes restants. Dans cette étape, nous utilisons un algorithme basé sur les distances de mots pour pouvoir identifier les doublons qui restent dans la base de données. Cet algorithme, développé par Bernard Kaddour¹⁰, permet de gérer les permutations de mots et les distances entre les mots. L'algorithme utilise notamment la distance de Levenshtein qui mesure la similarité entre deux chaînes de caractère. Elle est égale au nombre minimal de caractères qu'il faut supprimer, insérer ou remplacer pour passer d'une chaîne à une autre.

Pour chacun des algorithmes, il faut donner les champs qui serviront à la recherche de doublons (dans notre cas, le nom et la ville). Dans la mesure où l'algorithme traite les permutations d'ensemble de mots, il est préférable de refaire le traitement précédent mais en conservant les espaces cette fois-ci. Ensuite, il faut fixer un seuil d'acceptation qui permet à l'algorithme de décider à partir de quel moment les deux entrées dans la base de données correspondent au même agent. Ce seuil dépend de la qualité de la base de données. Il faut trouver un juste milieu entre des seuils trop bas qui conduirait à la création de beaucoup de singletons sauf en cas de correspondance parfaite et un seuil trop haut qui conduirait à l'effet inverse en attribuant à des agents différents des identifiants identiques.

Cet algorithme ne peut pas traiter la base de données en une seule fois. Pour optimiser la recherche de doublons, nous rentrons les données pays par pays. De plus, l'identifiant trouvé en étape 2 est donné en entrée à l'algorithme. A l'issue de cette étape, nous obtenons un identifiant plus robuste. Toutefois, celui-ci n'est pas encore parfait puisque pour éviter trop de défaut positif (attribuer le même identifiant à deux agents différents aux noms

¹⁰Université Jean Monnet, Master Web Intelligence.

proche comme par exemple Valence et Valencia), nous avons fixé les seuils d’acceptation assez bas. De ce fait, certaines entrées concernant le même agent se sont vues attribuer des identifiants différents et, au contraire, certains agents possèdent le même identifiant alors qu’en réalité il s’agit de deux agents distincts. C’est pourquoi dans un quatrième étape, nous procéderons à une vérification manuelle afin de corriger les dernières erreurs que l’on peut traiter.

2.4. Étape 4 : Vérification manuelle

Enfin, la dernière étape consiste à harmoniser les caractéristiques de chaque agent et à compléter les informations le concernant notamment en termes de localisation (vérifier notamment le cas des structures qui sont localisées dans les mêmes villes). Un agent peut participer à plusieurs projets dans le cadre du 6^{ème} PCRDT. Pour un même agent, il est possible que d’une participation à une autre certaines informations aient évoluées comme par exemple le type de l’agent. Il faut donc harmoniser les données portant sur cet agent pour qu’au final pour chacune de ces participations un même agent aient des caractéristiques identiques. Afin d’harmoniser les caractéristiques propres à chaque agent, si pour un même critère, il existe des divergences alors nous prendrons toujours en compte l’information la plus récente. Les différences d’une participation à une autre peuvent être dues à une évolution de l’agent mais aussi à des erreurs de saisie. Pour illustrer nos propos prenons l’exemple fictif suivant :

Année	Projet	Nom	Type 1	PME	Identifiant
2002	Projet A	FP1	IND	Oui	1
2006	Projet B	FP1	IND	Non	1

Dans cet exemple, l’entreprise “FP1” participe deux fois aux PCRDT en 2002 et en 2006. Entre ces deux participations, l’entreprise a changé et n’est plus une PME. Pour harmoniser la base de données, nous prenons en compte l’information la plus récente. Nous obtenons le résultat suivant :

Année	Projet	Nom	Type 1	PME	Identifiant
2002	Projet A	FP1	IND	Non	1
2006	Projet B	FP1	IND	Non	1

De cette manière, chaque identifiant au sein de la base renvoie à un seul agent qui possède un certain nombre de caractéristiques, ces dernières étant uniques sur l’ensemble du PCRDT. Le cas qui demande le plus grand nombre de traitement concerne les centres de recherche affiliés aux universités. Ces

derniers peuvent se déclarer en tant que centre de recherche mais aussi en tant qu'agent de l'enseignement supérieur. Le plus souvent lorsqu'une université participe à un programme cadre c'est un des laboratoires rattachés à l'université qui participe. Afin d'harmoniser la base de données, tous les laboratoires affiliés à une université sont considérés comme agent de l'enseignement supérieur. Dans la pratique et malgré la règle mis en place, il est parfois difficile de distinguer clairement entre centre de recherche et agent de l'enseignement supérieur. Enfin, la distinction entre PME et non PME est prise en compte seulement pour les entreprises.

Grâce à cette harmonisation, nous évitons les doubles comptages lorsque l'on s'intéresse notamment au nombre de PME qui participe aux PCRDT ou au nombre d'agents de chaque type. Cette harmonisation est nécessaire dans la mesure où le 6^{ème} PCRDT forme un seul réseau de collaborations sur la période 2002-2006. Il faut être conscient que cette harmonisation nécessaire entraîne un biais dans l'étude notamment au niveau des PME.

3. Les problèmes persistants

A l'issue de cas 4 étapes, un nombre important de corrections a été apporté à la base de données. Ces corrections permettent d'éviter un certain nombre de biais pendant l'étude comme par exemple un nombre d'agent plus important que ce qu'il n'est réellement. De plus, en harmonisant les caractéristiques de chaque agent, on évite aussi des doubles comptages lorsque l'on s'intéresse à des populations particulières. Toutefois, il persiste une certain nombre d'imperfections au sein de la base utilisée pour le thèse. En effet, plusieurs cas ne peuvent être traités à l'aide de requêtes ou d'algorithme. Les cas suivants ne peuvent pas être corrigés :

- les fusions acquisitions ;
- les changements de noms ;
- les changements d'adresse ;
- les filiales d'un groupe qui portent un nom différent ;
- les déménagements qui peuvent laisser penser qu'un groupe possède deux agents différents qui participent aux PCRDT alors qu'il s'agit en réalité du même agent qui a déménagé ;
- les agents plus particulièrement les entreprises qui déclarent le siège social au lieu de la filiale réellement impliquée dans le projet.

Les problèmes cités ci-dessus ne peuvent être gérés de manière automatique. Il faudrait pour chaque entrée de la base réaliser un traitement manuel en vérifiant les informations à l'aide d'internet.

La méthode employée ci-dessus est en cours d'amélioration.

Annexes A.4 : Robustesse des réseaux à la suppression des agents les plus centraux.

Agents supprimés	0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	Evolution globale
Diamètre	5	6	6	7	7	7	7	7	8	8	9	4
Évolution en %	-	20	0	16,7	0	0	0	0	14,3	0	12,5	80
Distance moyenne	2,42	2,706	2,821	2,961	3,057	3,242	3,435	3,636	3,758	3,89	4,194	1,8
Évolution en %	-	11,8	4,2	5,0	3,2	6,1	6,0	5,9	3,4	3,5	7,8	73,3
Nombre de composantes	1	2	2	3	3	4	5	6	8	10	10	9
Évolution en %	-	100,0	0,0	50,0	0,0	33,3	25,0	20,0	33,3	25,0	0,0	900
Grande composante	997	985	976	960	950	938	924	912	863	831	821	-176
Évolution en %	-	-1,2	-0,9	-1,6	-1,0	-1,3	-1,5	-1,3	-5,4	-3,7	-1,2	-17,7
Taille en %	100	99,8	99,9	99,3	99,3	99,0	98,6	98,4	94,1	91,6	91,5	
Deuxième composante	0	2	3	5	5	5	5	5	36	36	36	36
Évolution en %	-	-	50,0	66,7	0,0	0,0	0,0	0,0	620,0	0,0	0,0	-
Densité	3,11	2,6	2,4	2,16	2,01	1,87	1,76	1,66	2,57	1,49	1,44	-1,67
Évolution en %	-	-16,4	-7,7	-10,0	-6,9	-7,0	-5,9	-5,7	54,8	-42,0	-3,4	-53,7
Liens restants	30924	25694	22892	20200	18390	16784	15444	14226	13160	12262	11544	-19380
Évolution en %	-	-16,9	-10,9	-11,8	-9,0	-8,7	-8,0	-7,9	-7,5	-6,8	-5,9	-62,7
Agents restants	997	987	977	967	957	947	937	927	917	907	897	-10,0

Tableau : Évolution des principales propriétés générales du réseau Electronique-Microélectronique suite à la suppression des agents les plus connectés. La taille en % est évaluée par rapport à la population restante et non par rapport à la population de départ. La distance moyenne et le diamètre sont calculés sur la plus grande composante.

Annexes A.4 : Robustesse des réseaux à la suppression des agents les plus centraux.

Agents supprimés	0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	Évolution globale
Diamètre	5	5	5	6	6	6	6	6	8	8	8	3
Évolution en %	-	0	0	20	0	0	0	0	33,3	0	0	60
Distance moyenne	2,37	2,544	2,644	2,732	2,824	2,913	3	3,084	3,26	3,374	3,487	1,1
Évolution en %	-	7,3	3,9	3,3	3,4	3,2	3,0	2,8	5,7	3,5	3,3	47,1
Nombre de composantes	5	7	7	8	11	13	13	13	14	14	16	11
Évolution en %	-	40	0	14,3	37,5	18,2	0	0	7,7	0	14,3	220
Grande composante	1198	1176	1164	1145	1127	1110	1098	1086	1070	1058	1036	-162
Évolution en %	-	-1,8	-1,0	-1,6	-1,6	-1,5	-1,1	-1,1	-1,5	-1,1	-2,1	-13,5
Taille en %	98,9	98,1	98,1	97,4	96,9	96,5	96,5	96,4	96,1	96,0	95,0	
Deuxième composante	6	6	6	7	7	7	7	7	7	7	9	3
Évolution en %	-	0	0	16,7	0	0	0	0	0	0	28,6	50
Densité	3,59	3,08	2,76	2,49	2,27	2,06	1,91	1,79	1,68	1,6	1,52	-2,1
Évolution en %	-	-14,2	-10,4	-9,8	-8,8	-9,3	-7,3	-6,3	-6,1	-4,8	-5,0	-57,7
Liens restants	52548	44288	38856	34338	30642	27252	24686	22648	20882	19422	18094	-34454
Évolution en %	-	-15,7	-12,3	-11,6	-10,8	-11,1	-9,4	-8,3	-7,8	-7,0	-6,8	-65,6
Agents restants	1211	1199	1187	1175	1163	1150	1138	1126	1114	1102	1090	-10,0

Tableau : Évolution des principales propriétés générales du réseau Télécommunications suite à la suppression des agents les plus connectés. La taille en % est évaluée par rapport à la population restante et non par rapport à la population de départ. La distance moyenne et le diamètre sont calculés sur la plus grande composante.

Annexes A.5.1 : Identification des “hubs” dans le réseau Electronique-Microélectronique

Agents	Type 1	Nbr proj*	Norm**	Rank	$CD^B(i)$	$CD^B(i)'$	Rank	$CC^B(i)$	Rank	$CB^B(i)$	Rank	$CE^V(i)$	Rank	Av. rank
2677	res	45	26,95	2	426	42,77	1	63,00	1	16,97	1	43,53	2	1,25
13830	res	48	28,74	1	377	37,85	2	61,14	2	12,35	2	45,03	1	1,43
2297	res	31	18,56	3	293	29,42	3	57,51	3	6,39	3	36,14	3	3,00
2303	res	21	12,57	6	282	28,31	4	57,51	3	3,76	6	35,66	4	4,29
30006	he	24	14,37	4	273	27,41	5	56,88	5	5,49	5	29,04	5	4,76
12873	he	18	10,78	7	249	25,00	6	56,66	6	6,09	4	18,10	13	6,23
12411	ind	17	10,18	9	186	18,67	8	53,84	10	2,33	8	25,56	6	7,96
15454	ind	23	13,77	5	169	16,97	14	53,46	13	2,22	10	24,90	7	8,46
13730	he	11	6,59	15	202	20,28	7	54,55	7	2,29	9	19,18	10	8,87
2769	he	12	7,19	12	185	18,57	9	53,78	11	1,58	16	21,55	8	10,57
3356	he	11	6,59	15	185	18,57	9	54,19	8	3,06	7	9,11	42	10,65
23575	he	9	5,39	22	180	18,07	11	53,98	9	1,07	33	19,16	11	13,56
23074	ind	18	10,78	7	144	14,46	25	51,96	25	1,51	19	17,84	14	14,41
3125	res	12	7,19	12	172	17,27	13	53,21	16	2,19	11	10,27	34	14,57
2447	he	13	7,78	10	155	15,56	20	53,40	14	1,54	18	16,41	17	14,89
3311	he	9	5,39	22	167	16,77	15	53,55	12	0,98	36	19,25	9	14,95
100003	he	11	6,59	15	166	16,67	16	53,15	17	1,64	15	11,91	24	16,87
21974	he	13	7,78	10	135	13,55	26	52,78	19	1,10	32	17,13	16	17,55
23021	he	11	6,59	15	158	15,86	17	52,92	18	1,34	21	14,93	19	17,77
22048	he	7	4,19	34	176	17,67	12	53,35	15	1,41	20	10,02	35	19,38

Tableau : Classement des 40 agents les plus centraux du réseau Electronique-Microélectronique. Le rang de l’agent (Rank) est donné pour chaque indicateur. Av. Rank renvoie au rang moyen de l’agent selon les 5 indicateurs. (* Nombre de projets de l’agent ; ** Mesure normalisée du nombre de projets égale au rapport entre le nombre de projets de l’agent et le nombre total de projets financés dans le domaine technologique Electronique-Microélectronique.)

Agents	Type 1	Nbr proj	Norm	Rank	$CD^B(i)$	$CD^B(i)'$	Rank	$CC^B(i)$	Rank	$CB^B(i)$	Rank	$CE^V(i)$	Rank	Av. rank
21382	he	10	5,99	19	154	15,46	22	52,34	24	1,69	14	9,91	37	20,99
21433	ind	10	5,99	19	122	12,25	30	51,96	25	0,68	46	19,10	12	21,64
13629	res	9	5,39	22	158	15,86	17	52,45	22	1,55	17	8,83	45	21,67
23121	he	8	4,79	26	157	15,76	19	52,73	20	1,05	34	13,68	20	22,68
21894	he	6	3,59	41	155	15,56	20	52,39	23	1,33	22	10,56	29	25,28
22160	he	9	5,39	22	130	13,05	28	51,69	30	0,93	39	16,27	18	25,55
22189	he	12	7,19	12	95	9,54	54	50,56	46	1,14	30	10,43	32	26,57
13207	he	7	4,19	34	152	15,26	23	51,71	29	1,25	25	11,73	27	27,11
23056	he	10	5,99	19	110	11,04	39	51,63	31	1,33	23	9,47	39	27,83
15115	ind	8	4,79	26	112	11,24	36	50,40	48	0,36	78	17,35	15	30,02
5080	res	7	4,19	34	131	13,15	27	51,79	28	1,28	24	8,00	55	30,86
15189	he	7	4,19	34	119	11,95	32	51,88	27	1,21	27	9,17	41	31,42
2974	he	5	2,99	54	147	14,76	24	52,53	21	0,89	40	8,38	52	32,89
13247	res	8	4,79	26	106	10,64	42	50,25	52	0,66	48	11,91	23	34,29
21470	he	8	4,79	26	101	10,14	49	51,63	31	0,66	49	10,45	31	34,77
2106	he	8	4,79	26	108	10,84	41	51,21	38	1,23	26	5,72	76	35,51
14932	he	6	3,59	41	116	11,65	34	51,63	31	0,47	69	11,81	25	35,57
4649	ind	6	3,59	41	130	13,05	28	51,31	37	1,20	28	5,16	85	37,14
21790	he	5	2,99	54	119	11,95	32	51,63	31	0,40	73	11,07	28	38,04
22267	ind	8	4,79	26	99	9,94	53	50,84	41	0,70	44	9,49	38	38,24

Tableau bis : Classement des 40 agents les plus centraux du réseau Electronique-Microélectronique. Le rang de l'agent (Rank) est donné pour chaque indicateur. Av. Rank renvoie au rang moyen de l'agent selon les 5 indicateurs. (* Nombre de projets de l'agent ; ** Mesure normalisée du nombre de projets égale au rapport entre le nombre de projets de l'agent et le nombre total de projets financés dans le domaine technologique Electronique-Microélectronique.)

Agents	Nom	Type 1	Ville	Pays
2106	Budapest university technology and economics	he	Budapest	Hongrie
2297	CEA Commissariat a l'energie atomique	res	Paris	France
2303	CNRS centre national recherche scientifique delegation Paris Michel-Ange	res	Paris	France
2447	Rheinisch westfaelische technische hochschule aach	he	Aachen	Allemagne
2677	Fraunhofer gesellschaft für angewandte forschung	res	Munich	Allemagne
2769	Chalmers university technology school microtechnology and nanoscience	he	Goteborg	Suède
2974	Czech technical university in Prague	he	Prague	Rép. Tchèque
3125	Centre suisse d'electronique et microtechnique	res	Neuchatel	Suisse
3311	Eidgenossische technische hochschule Zurich	he	Zurich	Suisse
3356	Institute for multimedia communications universidad politecnica Valencia	he	Valencia	Espagne
4649	Centro Ricerche Fiat Societa Consortile per Azioni	ind	Orbasano	Italie
5080	Foundation for research and technology Hellas	res	Heraklion (Crete)	Grèce
12411	Infineon technologies AG	ind	Munich	Allemagne
12873	Kungliga tekniska högskolan royal institute technology	he	Stockholm	Suède
13207	Institute communication and computer studies technical university Athens	he	Athens	Grèce
13247	National centre for scientific research demokritos	res	Athens	Grèce
13629	Institute physics Polish academy sciences	res	Warsaw	Pologne
13730	Institute communication networks Vienna university technology	he	Vienna	Autriche
13830	IMEC Interuniversitair micoe electronica centrum vzw	res	Leuven	Belgique
14932	Linkoping university	he	Linkoping	Suède

Tableau : Informations complémentaires sur les 40 agents les plus centraux du réseau Electronique-Microélectronique.
Le tableau ne contient aucune PME.

Agents	Nom	Type 1	Ville	Pays
15115	Philips innovative technology solutions n v	ind	Haasrode	Belgique
15189	Lund university dept information technology	he	Lund	Suède
15454	Philips electronics b v	ind	Eindhoven	Pays-Bas
21382	Universidad politecnica Madrid	he	Madrid	Espagne
21433	ST Microelectronics SA	ind	Montrouge	France
21470	Eindhoven university technology	he	Eindhoven	Pays-Bas
21790	Universita di Bologna arces advanced research center for electronic systems e castro	he	Bologna	Italie
21894	University Stuttgart	he	Stuttgart	Allemagne
21974	Chancello masters and scholars university Cambridge	he	Cambridge	Royaume-Uni
22048	Technical university Catalonia	he	Barcelona	Espagne
22160	National university Ireland Cork university college Cork	he	Cork	Irlande
22189	Technical university Berlin	he	Berlin	Allemagne
22267	Thales SA	ind	Neuilly-sur-Seine	France
23021	Technical university Delft	he	Delft	Pays-Bas
23056	University Twente	he	Enschede	Pays-Bas
23074	ST Microelectronics srl	ind	Agratebrianza	Italie
23121	Warsaw university technology	he	Warsaw	Pologne
23575	Uppsala universitet	he	Uppsala	Suède
30006	Ecole polytechnique federale Lausanne	he	Lausanne	Suisse
100003	Katholieke universiteit Leuven	he	Leuven	Belgique

Tableau bis : Informations complémentaires sur les 40 agents les plus centraux du réseau Electronique-Microélectronique.
Le tableau ne contient aucune PME.

Annexes A.5.2 : Identification des “hubs” dans le réseau Télécommunications

Agents	Type 1	Nbr proj	Norm	Rank	$CD^B(i)$	$CD^B(i)'$	Rank	$CC^B(i)$	Rank	$CB^B(i)$	Rank	$CE^V(i)$	Rank	Av. rank
2677	res	56	29,95	1	529	44,19	1	63,74	1	14,01	1	32,66	2	1,11
21341	ind	49	26,20	2	502	41,94	2	62,87	2	8,58	2	38,40	1	1,67
2383	ind	35	18,72	3	366	30,58	3	58,30	3	3,05	6	29,82	3	3,33
12447	ind	32	17,11	4	348	29,07	4	57,94	4	3,23	4	25,40	7	4,38
23648	he	31	16,58	5	342	28,57	5	57,49	6	2,50	9	26,33	4	5,39
13207	he	28	14,97	6	342	28,57	5	56,84	7	3,16	5	25,91	6	5,71
22048	he	28	14,97	6	340	28,40	7	57,77	5	2,58	8	26,17	5	5,99
2106	he	23	12,30	8	295	24,64	9	56,54	8	2,21	15	19,67	9	9,28
11688	he	23	12,30	8	312	26,07	8	56,20	9	2,28	13	18,35	12	9,59
20219	ind	18	9,63	10	277	23,14	10	55,88	10	1,56	20	20,06	8	10,53
12135	he	17	9,09	12	272	22,72	11	55,37	12	2,23	14	17,17	15	12,64
23121	he	10	5,35	54	120	10,03	93	51,22	73	5,36	3	6,43	86	12,89
21093	ind	18	9,63	10	213	17,79	31	53,68	26	2,76	7	16,24	21	13,84
12873	he	15	8,02	22	271	22,64	12	55,78	11	1,24	27	18,59	11	14,38
30006	he	16	8,56	16	261	21,80	14	55,24	13	2,49	10	11,43	42	14,94
2447	he	16	8,56	16	254	21,22	16	55,09	14	1,14	31	18,03	14	16,66
22663	ind	17	9,09	12	256	21,39	15	53,41	30	2,34	12	7,46	67	17,76
20365	he	16	8,56	16	230	19,21	20	54,56	17	0,73	48	18,06	13	18,58
3311	he	16	8,56	16	252	21,05	17	54,68	16	1,49	23	14,43	25	18,71
13398	ind	17	9,09	12	229	19,13	22	53,58	27	1,50	22	14,59	23	19,63

Tableau : Classement des 40 agents les plus centraux du réseau Télécommunications. Le rang de l’agent (Rank) est donné pour chaque indicateur. Av. Rank renvoie au rang moyen de l’agent selon les 5 indicateurs. (* Nombre de projets de l’agent ; ** Mesure normalisée du nombre de projets égale au rapport entre le nombre de projets de l’agent et le nombre total de projets financés dans le domaine technologique Télécommunications.)

Agents	Type 1	Nbr proj	Norm	Rank	$CD^B(i)$	$CD^B(i)'$	Rank	$CC^B(i)$	Rank	$CB^B(i)$	Rank	$CE^V(i)$	Rank	Av. rank
3936	ind	17	9,09	12	197	16,46	37	52,66	43	0,50	65	18,60	10	20,08
21626	he	16	8,56	16	227	18,96	23	54,71	15	1,11	33	16,08	22	20,13
21382	he	14	7,49	23	242	20,22	18	53,56	29	2,39	11	7,05	76	21,05
11925	he	12	6,42	36	268	22,39	13	54,33	18	1,70	18	10,05	51	21,24
13100	he	16	8,56	16	224	18,71	27	54,16	21	0,52	60	16,33	18	22,79
5279	he	14	7,49	23	226	18,88	25	54,19	19	0,76	46	16,30	19	23,76
2264	ind	14	7,49	23	227	18,96	23	54,14	22	1,44	24	12,23	34	24,57
23330	he	14	7,49	23	232	19,38	19	53,13	34	1,68	19	10,46	47	25,07
21470	he	13	6,95	32	208	17,38	33	53,32	31	1,85	16	13,57	27	25,86
1810	res	13	6,95	32	221	18,46	28	54,19	19	1,24	28	13,52	28	26,17
15189	he	11	5,88	41	230	19,21	20	54,14	22	0,36	85	16,29	20	27,53
100003	he	14	7,49	23	218	18,21	29	52,92	39	1,56	21	7,14	73	30,32
3830	he	11	5,88	41	194	16,21	39	53,73	25	0,22	107	16,90	17	31,61
13243	he	11	5,88	41	215	17,96	30	53,99	24	0,48	69	14,46	24	32,14
13889	res	14	7,49	23	208	17,38	33	53,15	33	1,05	35	9,63	53	33,00
3629	ind	12	6,42	36	189	15,79	43	52,32	46	0,23	105	17,11	16	34,53
1309	res	13	6,95	32	196	16,37	38	53,58	27	0,81	42	11,87	38	34,55
13830	res	14	7,49	23	205	17,13	35	52,23	50	0,87	40	11,61	39	35,04
10024	he	11	5,88	41	190	15,87	41	53,01	36	0,63	52	13,15	29	38,38
21536	he	10	5,35	54	225	18,80	26	53,18	32	0,81	43	8,34	58	38,84

Tableau bis : Classement des 40 agents les plus centraux du réseau Télécommunications. Le rang de l'agent (Rank) est donné pour chaque indicateur. Av. Rank renvoie au rang moyen de l'agent selon les 5 indicateurs. (* Nombre de projets de l'agent ; ** Mesure normalisée du nombre de projets égale au rapport entre le nombre de projets de l'agent et le nombre total de projets financés dans le domaine technologique Télécommunications.)

Agents	Nom	Type 1	Ville	Pays
1309	Institut national recherche en informatique et automatique	res	Le Chesnay	France
1810	Technical research centre Finland	res	Espoo	Finlande
2106	Budapest university technology and economics	he	Budapest	Hongrie
2264	Alcatel space	ind	Paris	France
2383	France Telecom SA	ind	Paris	France
2447	Aachen university technology rwth Aachen	he	Aachen	Allemagne
2677	Fraunhofer gesellschaft zur foerderung angewandten forschung e v	res	Munich	Allemagne
3311	Swiss federal institute technology eth Zürich	he	Zurich	Suisse
3629	Ericsson ab	ind	Stockholm	Suède
3830	University science and technology	he	Cracow	Pologne
3936	Alcatel ag	ind	Stuttgart	Allemagne
5279	Technical university Denmark	he	Lyngby	Danemark
10024	Helsinki university technology	he	Espoo	Finlande
11688	Ecole nationale superieure telecommunications	he	Paris	France
11925	Forschungszentrum informatik an universität	he	Karlsruhe	Allemagne
12135	Politecnico di Torino dipartimento di elettronica	he	Torino	Italie
12447	Siemens ag	ind	Munich	Allemagne
12873	Kungliga tekniska högskolan	he	Stockholm	Suède
13100	Instituto superior tecnico	he	Lisboa	Portugal
13207	Institute communication and computer studies technical university Athens	he	Athens	Grèce

Tableau : Informations complémentaires sur les 40 agents les plus centraux du réseau Electronique-Microélectronique.
Le tableau ne contient aucune PME.

Agents	Nom	Type 1	Ville	Pays
13243	Athens university economics and business	he	Athens	Grèce
13398	Nokia corporation	ind	Espoo	Finlande
13830	IMEC Interuniversitair micoe electronica centrum vzw	res	Leuven	Belgique
13889	Institut eurecom	res	Sophia Antipolis	France
15189	Lund university dept information technology	he	Lund	Suède
20219	Telenor communication ii as	ind	Fonebu	Norvège
20365	Universidad Carlos III Madrid	he	Leganes	Espagne
21093	Telecom Italia	ind	Milan	Italie
21341	Telefonica investigacion desarrollo sociedad anonima	ind	Madrid	Espagne
21382	Universidad politecnica Madrid	he	Madrid	Espagne
21470	Eindhoven university technology	he	Eindhoven	Pays-Bas
21536	University Bristol UK	he	Bristol	Royaume-Uni
21626	University College London	he	London	Royaume-Uni
22048	Technical university Catalonia	he	Barcelona	Espagne
22663	Thales	ind	Colombes	France
23121	Warsaw univeristy technology	he	Warsaw	Pologne
23330	Dresden university technology	he	Dresden	Allemagne
23648	University Surrey	he	Guildford	Royaume-Uni
30006	Swiss federal institute technology Lausanne	he	Lausanne	Suisse
100003	Katholiek universiteit Leuven	he	Leuven	Belgique

Tableau bis : Informations complémentaires sur les 40 agents les plus centraux du réseau Electronique-Microélectronique.
Le tableau ne contient aucune PME.

Annexes A.6 : Identification des “hubs” dans le réseau agrégé

Agents	Type 1	Nbr proj	Norm	Rank	$CD^B(i)$	$CD^B(i)'$	Rank	$CC^B(i)$	Rank	$CB^B(i)$	Rank	$CE^V(i)$	Rank	Av. rank
2677	res	101	28,53	1	854	44,16	1	63,85	1	15,84	1	37,80	1	1,00
21341	ind	54	15,25	3	566	29,27	2	58,03	2	4,58	3	33,96	2	2,31
13830	res	62	17,51	2	524	27,09	3	57,07	3	5,58	2	20,88	9	2,81
30006	he	40	11,30	5	483	24,97	4	56,50	4	4,33	4	17,15	16	4,94
22048	he	35	9,89	10	458	23,68	5	56,14	5	2,07	11	25,00	4	5,95
2383	ind	38	10,73	6	397	20,53	10	54,99	10	1,60	16	26,62	3	6,56
2297	res	45	12,71	4	431	22,29	8	55,27	9	3,15	6	17,93	12	6,79
13207	he	35	9,89	10	441	22,80	7	55,64	7	2,27	9	24,99	5	7,18
12873	he	33	9,32	12	443	22,91	6	56,11	6	2,93	7	21,24	8	7,30
23648	he	36	10,17	8	429	22,18	9	55,34	8	1,79	15	24,41	6	8,41
12447	ind	37	10,45	7	379	19,60	12	54,87	11	1,97	14	21,39	7	9,41
2106	he	31	8,76	13	373	19,29	13	54,76	12	2,03	12	18,19	11	12,15
2303	res	30	8,47	14	393	20,32	11	54,57	15	2,25	10	11,62	41	14,15
2447	he	29	8,19	15	360	18,61	16	54,68	13	1,36	26	19,45	10	14,51
3311	he	25	7,06	16	372	19,23	14	54,68	13	1,48	21	17,62	13	14,91
21382	he	24	6,78	19	362	18,72	15	54,28	16	2,43	8	9,64	51	15,32
15454	ind	36	10,17	8	318	16,44	21	53,41	23	1,55	19	12,76	33	16,72
23121	he	18	5,08	36	258	13,34	41	52,64	35	3,85	5	9,10	61	16,83
100003	he	25	7,06	16	349	18,05	17	53,99	18	1,98	13	9,67	50	18,26
12135	he	21	5,93	26	335	17,32	18	54,17	17	1,46	23	17,20	14	18,67

Tableau : Classement des 40 agents les plus centraux du réseau agrégé. Le rang de l’agent (Rank) est donné pour chaque indicateur. Av. Rank renvoie au rang moyen de l’agent selon les 5 indicateurs. (* Nombre de projets de l’agent ; ** Mesure normalisée du nombre de projets égale au rapport entre le nombre de projets de l’agent et le nombre total de projets financés dans le domaine technologique Télécommunications.)

Annexes A.6 : Identification des “hubs” dans le réseau agrégé

Agents	Type 1	Nbr proj	Norm	Rank	$CD^B(i)$	$CD^B(i)'$	Rank	$CC^B(i)$	Rank	$CB^B(i)$	Rank	$CE^V(i)$	Rank	Av. rank
11688	he	24	6,78	19	333	17,22	19	53,63	20	1,13	34	16,35	19	21,07
3356	he	22	6,21	25	320	16,55	20	53,98	19	1,37	25	14,54	24	22,29
15189	he	18	5,08	36	303	15,67	22	53,45	21	0,65	56	17,18	15	24,35
22663	ind	25	7,06	16	299	15,46	24	52,47	41	1,56	18	7,45	79	25,41
1810	res	21	5,93	26	303	15,67	22	53,43	22	1,04	37	14,12	27	25,85
21470	he	21	5,93	26	273	14,12	36	52,97	29	1,34	27	14,96	22	27,29
13398	ind	24	6,78	19	278	14,37	34	52,51	40	1,30	28	13,98	30	28,39
5279	he	20	5,65	29	271	14,01	37	53,02	28	0,65	57	16,67	17	28,80
12411	ind	24	6,78	19	279	14,43	33	52,54	38	1,48	22	9,57	56	28,98
21894	he	16	4,52	47	292	15,10	27	53,28	24	0,78	44	14,65	23	30,09
20219	ind	18	5,08	36	277	14,32	35	52,88	31	0,69	52	16,42	18	30,60
23074	ind	24	6,78	19	283	14,63	31	52,88	31	1,05	36	9,46	57	30,78
21093	ind	20	5,65	29	221	11,43	57	51,67	58	1,59	17	13,36	32	31,38
1309	res	19	5,37	32	268	13,86	38	53,15	26	1,30	29	12,11	35	31,43
2264	ind	19	5,37	32	281	14,53	32	53,13	27	1,06	35	12,22	34	31,74
13243	he	16	4,52	47	295	15,25	25	53,20	25	0,62	60	14,24	26	31,97
2769	he	18	5,08	36	284	14,68	30	52,94	30	0,78	43	11,81	36	34,37
21626	he	20	5,65	29	244	12,62	45	52,61	36	0,66	55	14,01	28	36,13
23330	he	19	5,37	32	288	14,89	28	51,92	52	1,18	33	9,99	46	36,17
20365	he	17	4,80	42	246	12,72	44	52,60	37	0,37	83	15,98	20	36,87

Tableau bis : Classement des 40 agents les plus centraux du réseau agrégé. Le rang de l’agent (Rank) est donné pour chaque indicateur. Av. Rank renvoie au rang moyen de l’agent selon les 5 indicateurs. (* Nombre de projets de l’agent ; ** Mesure normalisée du nombre de projets égale au rapport entre le nombre de projets de l’agent et le nombre total de projets financés dans le domaine technologique Télécommunications.)

Agents	Nom	Type 1	Ville	Pays
1309	Institut national recherche en informatique et automatique	res	Le Chesnay	France
1810	Valition teknillinen tutkimuskeskus vtt technical	res	Espoo	Finlande
2106	Budapest university technology and economics	he	Budapest	Hongrie
2264	Alcatel cit	ind	Paris	France
2297	CEA Commissariat a l'énergie atomique	res	Paris	France
2303	CNRS Centre National Recherche Scientifique	res	Paris	France
2383	France Telecom	ind	Paris	France
2447	Aachen university technology rwth Aachen	he	Aachen	Allemagne
2677	Fraunhofer gesellschaft zur foerderung angewandten forschung e v	res	Munich	Allemagne
2769	Chalmers university technology school microtechnology and nanoscience	he	Goteborg	Suède
3311	Swiss federal institute technology eth Zurich	he	Zurich	Suisse
3356	Universidad politecnica Valencia	he	Valencia	Espagne
5279	Technical university Denmark	he	Lyngby	Danemark
11688	Ecole nationale superieure telecommunications	he	Paris	France
12135	Politecnico di Torino dipartimento di elettronica	he	Torino	Italie
12411	Infineon technologies	ind	Munich	Allemagne
12447	Siemens ag	ind	Munich	Allemagne
12873	Kungliga tekniska högskolan royal institute technology	he	Stockholm	Suède
13207	Institute communication and computer studies national technical university Athens	he	Athens	Grèce
13243	Athens university economics and business	he	Athens	Grèce

Tableau : Informations complémentaires sur les 40 agents les plus centraux du réseau agrégé. Le tableau ne contient aucune PME.

Agents	Nom	Type 1	Ville	Pays
13398	Nokia corporation	ind	Espoo	Finlande
13830	Interuniversitair micoe electronica centrum vzw	res	Leuven	Belgique
15189	Department communication systems	he	Lund	Suède
15454	Philips	ind	Eindhoven	Pays-Bas
20219	Telenor	ind	Fornebu	Norvège
20365	Universidad Carlos III madrid	he	Madrid	Espagne
21093	Telecom Italia	ind	Milan	Italie
21341	Telefonica investigacion y desarrollo	ind	Madrid	Espagne
21382	Universidad politecnica Madrid	he	Madrid	Espagne
21470	Eindhoven university technology	he	Eindhoven	Pays-Bas
21626	University college of London	he	London	Royaume-uni
21894	Universität Stuttgart	he	Stuttgart	Allemagne
22048	Technical university Catalonia	he	Barcelona	Espagne
22663	Thales communication	ind	Colombes	France
23074	ST Microelectronics srl	ind	Agrate-Brianza	Italie
23121	Warsaw univeristy technology	he	Warsaw	Pologne
23330	Dresden university technology	he	Dresden	Allemagne
23648	University Surrey	he	Guildford	Royaume-uni
30006	Ecole polytechnique federale Lausanne	he	Lausanne	Suisse
100003	Katholieke univiersiteit Leuven	he	Leuven	Belgique

Tableau bis : Informations complémentaires sur les 40 agents les plus centraux du réseau agrégé. Le tableau ne contient aucune PME.

Annexes A.7.1 : Robustesse des réseaux à la suppression des agents les plus centraux.

Agents supprimés	0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	Evolution globale
Diamètre	8	9	10	10	10	13	13	14	14	14	14	6
Évolution en %	-	12,5	11,1	0,0	0,0	30,0	0,0	7,7	0,0	0,0	0,0	75
Distance moyenne	3,158	3,651	4,097	4,292	4,612	5,544	5,804	6,114	6,321	6,417	6,51	3,352
Évolution en %	-	15,6	12,2	4,8	7,5	20,2	4,7	5,3	3,4	1,5	1,4	106,1
Nombre de composantes	19	23	26	28	28	31	33	34	35	35	35	16
Évolution en %	-	21,1	13,0	7,7	0,0	10,7	6,5	3,0	2,9	0,0	0,0	84,2
Grande composante	327	307	295	287	283	258	237	234	229	228	224	-103
Évolution en %	-	-6,1	-3,9	-2,7	-1,4	-8,8	-8,1	-1,3	-2,1	-0,4	-1,8	-31,5
Taille en %	86,3	81,9	79,5	78,0	77,7	71,7	66,6	66,5	65,6	66,1	65,7	-
Deuxième composante	6	6	6	6	6	13	13	13	13	13	13	7
Évolution en %	-	0,0	0,0	0,0	0,0	116,7	0,0	0,0	0,0	0,0	0,0	116,7
Densité	2,26	1,93	1,75	1,65	1,55	1,46	1,4	1,37	1,35	1,34	1,35	-0,91
Évolution en %	-	-14,6	-9,3	-5,7	-6,1	-5,8	-4,1	-2,1	-1,5	-0,7	0,7	-40,3
Liens restants	3238	2700	2404	2234	2050	1892	1774	1694	1640	1596	1568	-1670
Évolution en %	-	-16,6	-11,0	-7,1	-8,2	-7,7	-6,2	-4,5	-3,2	-2,7	-1,8	-51,6
Agents restants	379	375	371	368	364	360	356	352	349	345	341	-10,0

Tableau : Évolution des principales propriétés générales du réseau FP5 nanotechnologies suite à la suppression des agents les plus connectés. La taille en % est évaluée par rapport à la population restante et non par rapport à la population de départ. La distance moyenne et le diamètre sont calculés sur la plus grande composante.

Annexes A.7.2 : Robustesse des réseaux à la suppression des agents les plus centraux.

Agents supprimés	0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	Evolution globale
Diamètre	5	6	7	7	7	7	7	7	8	8	8	3
Évolution en %	-	20,0	16,7	0,0	0,0	0,0	0,0	0,0	14,3	0,0	0,0	60
Distance moyenne	2,357	2,725	3	3,172	3,36	3,495	3,58	3,663	3,861	3,924	3,929	1,572
Évolution en %	-	15,6	10,1	5,7	5,9	4,0	2,4	2,3	5,4	1,6	0,1	66,7
Nombre de composantes	1	2	2	2	2	3	5	7	7	7	7	6
Évolution en %	-	100,0	0,0	0,0	0,0	50,0	66,7	40,0	0,0	0,0	0,0	600,0
Grande composante	522	512	507	501	496	480	425	418	416	414	414	-108
Évolution en %	-	-1,9	-1,0	-1,2	-1,0	-3,2	-11,5	-1,6	-0,5	-0,5	0,0	-20,7
Taille en %	100,0	99,0	99,0	99,0	99,0	96,8	86,6	86,2	86,7	87,2	88,1	-
Deuxième composante	0	5	5	5	5	11	45	40	37	34	29	29
Évolution en %	-	-	0,0	0,0	0,0	120,0	309,1	-11,1	-7,5	-8,1	-14,7	-
Densité	4,85	4,23	3,9	3,62	3,45	3,29	3,16	3,03	2,93	2,84	2,76	-2,09
Évolution en %	-	-12,8	-7,8	-7,2	-4,7	-4,6	-4,0	-4,1	-3,3	-3,1	-2,8	-43,1
Liens restants	13182	11288	10214	9260	8632	8088	7604	7124	6738	6386	6076	-7106
Évolution en %	-	-14,4	-9,5	-9,3	-6,8	-6,3	-6,0	-6,3	-5,4	-5,2	-4,9	-53,9
Agents restants	522	517	512	506	501	496	491	485	480	475	470	-10,0

Tableau : Évolution des principales propriétés générales du réseau FP6 nanotechnologies suite à la suppression des agents les plus connectés. La taille en % est évaluée par rapport à la population restante et non par rapport à la population de départ. La distance moyenne et le diamètre sont calculés sur la plus grande composante.

Annexes A.8.1 : Identification des “hubs” dans le réseau nanotechnologies du 5^{ème} PCRD

Agents	Type 1	Nbr proj	Norm	Rank	$CD^B(i)$	$CD^B(i)'$	Rank	$CC^B(i)$	Rank	$CB^B(i)$	Rank	$CE^V(i)$	Rank	Av. rank
13830	res	23	19,33	1	86	26,38	1	48,73	2	19,55	1	47,90	2	1,25
12411	ind	14	11,76	2	75	23,01	2	49,10	1	14,30	2	50,09	1	1,43
4383	res	13	10,92	3	59	18,10	3	47,66	3	9,70	3	36,19	4	3,16
23083	ind	10	8,40	4	55	16,87	4	45,98	4	5,44	7	45,86	3	4,08
3125	res	6	5,04	8	44	13,50	5	45,34	5	8,00	4	23,06	8	5,56
15454	ind	6	5,04	8	44	13,50	5	43,76	8	2,07	25	31,08	5	7,25
22160	he	9	7,56	5	39	11,96	7	43,94	7	6,08	6	12,20	29	7,28
30006	he	6	5,04	8	37	11,35	8	44,97	6	4,71	8	20,09	11	7,90
100070	res	8	6,72	6	30	9,20	13	41,90	13	1,22	34	25,42	6	9,68
30005	ind	5	4,20	13	30	9,20	13	39,76	25	7,62	5	8,04	44	12,00
3591	res	5	4,20	13	33	10,12	10	43,24	10	3,12	15	15,44	19	12,62
23074	ind	5	4,20	13	26	7,98	17	43,35	9	3,03	17	17,01	14	13,26
22048	he	4	3,36	17	36	11,04	9	40,75	19	3,86	12	16,32	15	13,42
100078	res	7	5,88	7	26	7,98	17	39,32	27	4,68	9	5,67	60	13,64
12135	he	5	4,20	13	31	9,51	12	41,85	14	1,25	33	20,00	12	14,48
13993	he	3	2,52	26	33	10,12	10	41,32	16	1,36	31	20,97	10	15,01
100026	ind	4	3,36	17	21	6,44	25	41,69	15	0,37	55	24,59	7	15,31
23765	he	2	1,68	38	25	7,67	20	42,17	12	0,28	65	22,72	9	17,47
100020	ind	4	3,36	17	22	6,75	23	42,39	11	1,02	35	10,20	33	19,83
13247	res	4	3,36	17	27	8,28	15	40,15	21	2,11	24	9,08	40	20,85

Tableau : Classement des 20 agents les plus centraux du réseau nanotechnologies du 5^{ème} PCRD. Le rang de l'agent (Rank) est donné pour chaque indicateur. Av. Rank renvoie au rang moyen de l'agent selon les 5 indicateurs. (* Nombre de projets de l'agent ; ** Mesure normalisée du nombre de projets égale au rapport entre le nombre de projets de l'agent et le nombre total de projets financés dans le domaine technologique Télécommunications.)

Agents	Nom	Type 1	Ville	Pays
3125	Centre Suisse d'électronique et microtechnique	res	Neuchatel	Suisse
3591	CNRS Centre National Recherche Scientifique	res	Toulouse	France
4383	CEA Commissariat a l'énergie atomique	res	Grenoble	France
12135	Politecnico di Torino dipartimento di elettronica cercom	he	Torino	Italie
12411	Infineon technologies	ind	Munich	Allemagne
13247	National center for scientific research demokritos	res	Athens	Grèce
13830	IMEC Interuniversitair micoe electronica centrum vzw	res	Leuven	Belgique
13993	Institut national polytechnique Grenoble	he	Grenoble	France
15454	Philips electronics b v	ind	Eindhoven	Pays-Bas
22048	Technical university Catalonia	he	Barcelona	Espagne
22160	National university Ireland Cork	he	Cork	Irlande
23074	ST Microelectronics	ind	Agrate-Brianza	Italie
23083	ST Microelectronics	ind	Crolles	France
23765	Universita degli studi di Udine	he	Udine	Italie
30005	AMI semiconductor Belgium bvba	ind	Oudenaarde	Belgique
30006	Ecole polytechnique federale Lausanne	he	Lausanne	Suisse
100020	Alcatel	ind	Oudenaarde	Belgique
100026	Philips NV	ind	Leuven	Belgique
100070	Fraunhofer Gesellschaft zur Foerderung Der Angewandten Forschung E.V.	res	Erlangen	Allemagne
100078	Fraunhofer Gesellschaft zur Foerderung Der Angewandten Forschung E.V.	res	Itzehoe	Allemagne

Tableau : Informations complémentaires sur les 20 agents les plus centraux du réseau nanotechnologies du 5^{ème} PCRD. Le tableau ne contient aucune PME.

Annexes A.8.2 : Identification des “hubs” dans le réseau nanotechnologies du 6^{ème} PCRD

Agents	Type 1	Nbr proj	Norm	Rank	$CD^B(i)$	$CD^B(i)'$	Rank	$CC^B(i)$	Rank	$CB^B(i)$	Rank	$CE^V(i)$	Rank	Av. rank
2677	res	24	32,43	2	249	47,79	1	64,88	1	21,15	1	45,85	1	1,11
13830	res	27	36,49	1	206	39,54	2	61,15	2	13,60	2	43,87	2	1,67
2297	res	19	25,68	3	205	39,35	3	61,01	3	12,75	3	41,34	3	3,00
3125	res	9	12,16	8	153	29,37	4	57,57	4	8,10	4	20,31	8	5,00
30006	he	12	16,22	6	143	27,45	5	56,69	5	6,84	5	26,44	7	5,50
12411	ind	10	13,51	7	117	22,46	6	55,72	6	3,53	8	28,97	4	5,87
23074	ind	13	17,57	4	115	22,07	7	54,55	9	3,75	7	27,09	6	6,15
15454	ind	13	17,57	4	104	19,96	10	55,02	8	3,07	10	28,29	5	6,45
13629	res	6	8,11	10	115	22,07	7	55,60	7	5,12	6	13,78	17	8,18
22160	he	6	8,11	10	95	18,23	12	53,77	12	2,38	12	17,76	10	11,11
13207	he	4	5,41	21	97	18,62	11	54,33	10	3,18	9	14,52	15	12,01
22059	he	4	5,41	21	115	22,07	7	53,55	13	2,66	11	13,14	19	12,17
21433	ind	6	8,11	10	77	14,78	17	52,84	15	1,35	22	19,13	9	13,09
2303	res	6	8,11	10	69	13,24	23	52,15	19	0,69	37	17,29	11	15,92
23648	he	3	4,05	37	88	16,89	14	53,82	11	1,61	19	14,22	16	16,42
22758	ind	5	6,76	15	83	15,93	15	53,00	14	1,50	21	12,95	20	16,54
23121	he	5	6,76	15	75	14,40	18	51,64	20	1,20	26	15,25	12	17,01
100003	he	7	9,46	9	72	13,82	21	50,83	29	1,69	17	11,28	29	17,45
22048	he	3	4,05	37	90	17,27	13	52,63	16	1,34	23	12,12	26	20,13
3984	he	5	6,76	15	70	13,44	22	52,20	18	1,18	27	9,237535	37	21,58

Tableau : Classement des 20 agents les plus centraux du réseau nanotechnologies du 6^{ème} PCRD. Le rang de l'agent (Rank) est donné pour chaque indicateur. Av. Rank renvoie au rang moyen de l'agent selon les 5 indicateurs. (* Nombre de projets de l'agent ; ** Mesure normalisée du nombre de projets égale au rapport entre le nombre de projets de l'agent et le nombre total de projets financés dans le domaine technologique Télécommunications.)

Agents	Nom	Type 1	Ville	Pays
2297	CEA Commissariat a l'énergie atomique	res	Paris	France
2303	CNRS Centre National Recherche Scientifique	res	Paris	France
2677	Fraunhofer Gesellschaft Zur Förderung Angewandten Forschung e v	res	Munich	Allemagne
3125	Centre Suisse d'électronique et microtechnique	res	Neuchatel	Suisse
3984	Albert Ludwigs universität Freiburg	he	Freiburg	Allemagne
12411	Infineon technologies	ind	Munich	Allemagne
13207	National technical university Athens	he	Athens	Grèce
13629	Institute electron technology	res	Warsaw	Pologne
13830	IMEC Interuniversitair micoe electronica centrum vzw	res	Leuven	Belgique
15454	Philips electronics BV	ind	Eindhoven	Pays-Bas
21433	ST Microelectronics	ind	Paris	France
22048	Technical university Catalonia	he	Barcelona	Espagne
22059	Universitat Barcelona	he	Barcelona	Espagne
22160	National university Ireland Ccork	he	Cork	Irlande
22758	VDI VDE innovation technik gmbh	ind	Teltow	Allemagne
23074	ST Microelectronics srl	ind	Agrate-Brianza	Italie
23121	Politechnika Warszawska Warsaw university tech	he	Warsaw	Pologne
23648	University Surrey	he	Guildford	Royaume-Uni
30006	Ecole polytechnique federal Lausanne	he	Lausanne	Suisse
100003	Katholieke univiersiteit Leuven	he	Leuven	Belgique

Tableau : Informations complémentaires sur les 20 agents les plus centraux du réseau nanotechnologies du 6^{ème} PCRD. Le tableau ne contient aucune PME.

Annexes A.9 : Indicateurs de participation des pays.

Pays	Nombre d'agents	Répartition en %	Nombre de projets	Répartition en %	Participation Moyenne
Allemagne	330	16,99	917	18,53	2,78
Autriche	56	2,88	121	2,45	2,16
Belgique	60	3,09	214	4,32	3,57
Danemark	35	1,80	83	1,68	2,37
Espagne	125	6,44	380	7,68	3,04
Finlande	47	2,42	131	2,65	2,79
France	277	14,26	696	14,07	2,51
Grèce	59	3,04	197	3,98	3,34
Irlande	24	1,24	46	0,93	1,92
Italie	188	9,68	447	9,03	2,38
Luxembourg	3	0,15	5	0,10	1,67
Pays-Bas	58	2,99	184	3,72	3,17
Portugal	26	1,34	95	1,92	3,65
Royaume-Uni	227	11,69	528	10,67	2,33
Suède	62	3,19	188	3,80	3,03
UE15*	1577	81,20	4232	85,53	2,68
Bulgarie	5	0,26	5	0,10	1,00
Chypre	4	0,21	9	0,18	2,25
Estonie	7	0,36	8	0,16	1,14
Hongrie	24	1,24	59	1,19	2,46
Lettonie	6	0,31	7	0,14	1,17
Lituanie	14	0,72	15	0,30	1,07
Malte	2	0,10	2	0,04	1,00
Pologne	36	1,85	91	1,84	2,53
République Tchèque	17	0,88	28	0,57	1,65
Roumanie	14	0,72	25	0,51	1,79
Slovaquie	9	0,46	9	0,18	1,00
Slovenie	7	0,36	16	0,32	2,29
NEM**	145	7,47	274	5,54	1,89
UE27***	1722	88,67	4506	91,07	2,62

(*UE15 = Union Européenne à 15 membres ; **NEM = Nouveaux États Membres ; ***UE27 = Union Européenne à 27 membres)

Tableau : Indicateurs de participations de chaque pays au réseau dans le réseau agrégé.

Pays	Nombre d'agents	Répartition en %	Nombre de projets	Répartition en %	Participation Moyenne
Ukraine	1	0,05	1	0,02	1,00
Islande	1	0,05	1	0,02	1,00
Norvège	26	1,34	67	1,35	2,58
Suisse	64	3,30	184	3,72	2,88
Argentine	1	0,05	1	0,02	1,00
Australie	4	0,21	7	0,14	1,75
Biélorussie	1	0,05	2	0,04	2,00
Brésil	4	0,21	4	0,08	1,00
Canada	5	0,26	12	0,24	2,40
Chine	9	0,46	14	0,28	1,56
Corée du Sud	5	0,26	7	0,14	1,40
Croatie	1	0,05	4	0,08	4,00
Etats-Unis	11	0,57	11	0,22	1,00
Inde	1	0,05	3	0,06	3,00
Israël	49	2,52	71	1,43	1,45
Japon	7	0,36	7	0,14	1,00
Liechtenstein	1	0,05	1	0,02	1,00
Russie	11	0,57	13	0,26	1,18
Serbie	1	0,05	1	0,02	1,00
Singapour	3	0,15	11	0,22	3,67
Taïwan	1	0,05	1	0,02	1,00
Turquie	13	0,67	19	0,38	1,46
Total/Moyenne	1942	100	4948	100	98,35

Tableau bis : Indicateurs de participations de chaque pays au réseau dans le réseau agrégé.

Pays	Nombre d'agents	Répartition en %	Nombre de projets	Répartition en %	Participation Moyenne
Allemagne	197	19,76	409	20,77	2,08
Autriche	28	2,81	44	2,23	1,57
Belgique	33	3,31	115	5,84	3,48
Danemark	19	1,91	32	1,63	1,68
Espagne	52	5,22	109	5,54	2,10
Finlande	27	2,71	45	2,29	1,67
France	154	15,45	299	15,19	1,94
Grèce	20	2,01	49	2,49	2,45
Irlande	13	1,30	24	1,22	1,85
Italie	79	7,92	155	7,87	1,96
Luxembourg	0	0,00	0	0,00	-
Pays-Bas	31	3,11	85	4,32	2,74
Portugal	9	0,90	14	0,71	1,56
Royaume-Uni	133	13,34	221	11,22	1,66
Suède	37	3,71	91	4,62	2,46
UE15*	832	83,45	1692	85,93	2,03
Bulgarie	1	0,10	1	0,05	1,00
Chypre	1	0,10	1	0,05	1,00
Estonie	2	0,20	2	0,10	1,00
Hongrie	11	1,10	20	1,02	1,82
Lettonie	1	0,10	1	0,05	1,00
Lituanie	7	0,70	7	0,36	1,00
Malte	1	0,10	1	0,05	1,00
Pologne	21	2,11	41	2,08	1,95
République Tchèque	7	0,70	15	0,76	2,14
Roumanie	7	0,70	12	0,61	1,71
Slovaquie	6	0,60	6	0,30	1,00
Slovenie	1	0,10	1	0,05	1,00
NEM**	66	6,62	108	5,49	1,64
UE27***	898	90,07	1800	91,42	2,00

(*UE15 = Union Européenne à 15 membres; **NEM = Nouveaux États Membres; ***UE27 = Union Européenne à 27 membres)

Tableau : Indicateurs de participations de chaque pays au réseau dans le réseau Electronique-Microélectronique.

Pays	Nombre d'agents	Répartition en %	Nombre de projets	Répartition en %	Participation Moyenne
Ukraine	1	0,10	1	0,05	1,00
Islande	0	0,00	0	0,00	-
Norvège	9	0,90	15	0,76	1,67
Suisse	38	3,81	92	4,67	2,42
Argentine	0	0,00	0	0,00	-
Australie	4	0,40	4	0,20	1,00
Biélorussie	1	0,10	2	0,10	2,00
Brésil	0	0,00	0	0,00	-
Canada	1	0,10	2	0,10	2,00
Chine	1	0,10	1	0,05	1,00
Corée du Sud	2	0,20	2	0,10	1,00
Croatie	1	0,10	1	0,05	1,00
Etats-Unis	4	0,40	4	0,20	1,00
Inde	0	0,00	0	0,00	-
Israël	24	2,41	29	1,47	1,21
Japon	0	0,00	0	0,00	-
Liechtenstein	1	0,10	1	0,05	1,00
Russie	5	0,50	6	0,30	1,20
Serbie	0	0,00	0	0,00	-
Singapour	1	0,10	1	0,05	1,00
Taïwan	0	0,00	0	0,00	-
Turquie	6	0,60	8	0,41	1,33
Total/Moyenne	997	100	1969	100	1,97

Tableau bis : Indicateurs de participations de chaque pays au réseau dans le réseau Electronique-Microélectronique.

Pays	Nombre d'agents	Répartition en %	Nombre de projets	Répartition en %	Participation Moyenne
Allemagne	180	14,86	508	17,05	2,82
Autriche	36	2,97	77	2,58	2,14
Belgique	36	2,97	99	3,32	2,75
Danemark	23	1,90	51	1,71	2,22
Espagne	92	7,60	271	9,10	2,95
Finlande	27	2,23	86	2,89	3,19
France	160	13,21	397	13,33	2,48
Grèce	52	4,29	148	4,97	2,85
Irlande	14	1,16	22	0,74	1,57
Italie	135	11,15	292	9,80	2,16
Luxembourg	3	0,25	5	0,17	1,67
Pays-Bas	33	2,73	99	3,32	3,00
Portugal	23	1,90	81	2,72	3,52
Royaume-Uni	128	10,57	307	10,31	2,40
Suède	34	2,81	97	3,26	2,85
UE15*	976	80,59	2540	85,26	2,60
Bulgarie	4	0,33	4	0,13	1,00
Chypre	4	0,33	8	0,27	2,00
Estonie	6	0,50	6	0,20	1,00
Hongrie	15	1,24	39	1,31	2,60
Lettonie	5	0,41	6	0,20	1,20
Lituanie	8	0,66	8	0,27	1,00
Malte	1	0,08	1	0,03	1,00
Pologne	21	1,73	50	1,68	2,38
République Tchèque	12	0,99	13	0,44	1,08
Roumanie	8	0,66	13	0,44	1,63
Slovaquie	3	0,25	3	0,10	1,00
Slovénie	7	0,58	15	0,50	2,14
NEM**	94	7,76	166	5,57	1,77
UE27***	1070	88,36	2706	90,84	2,53

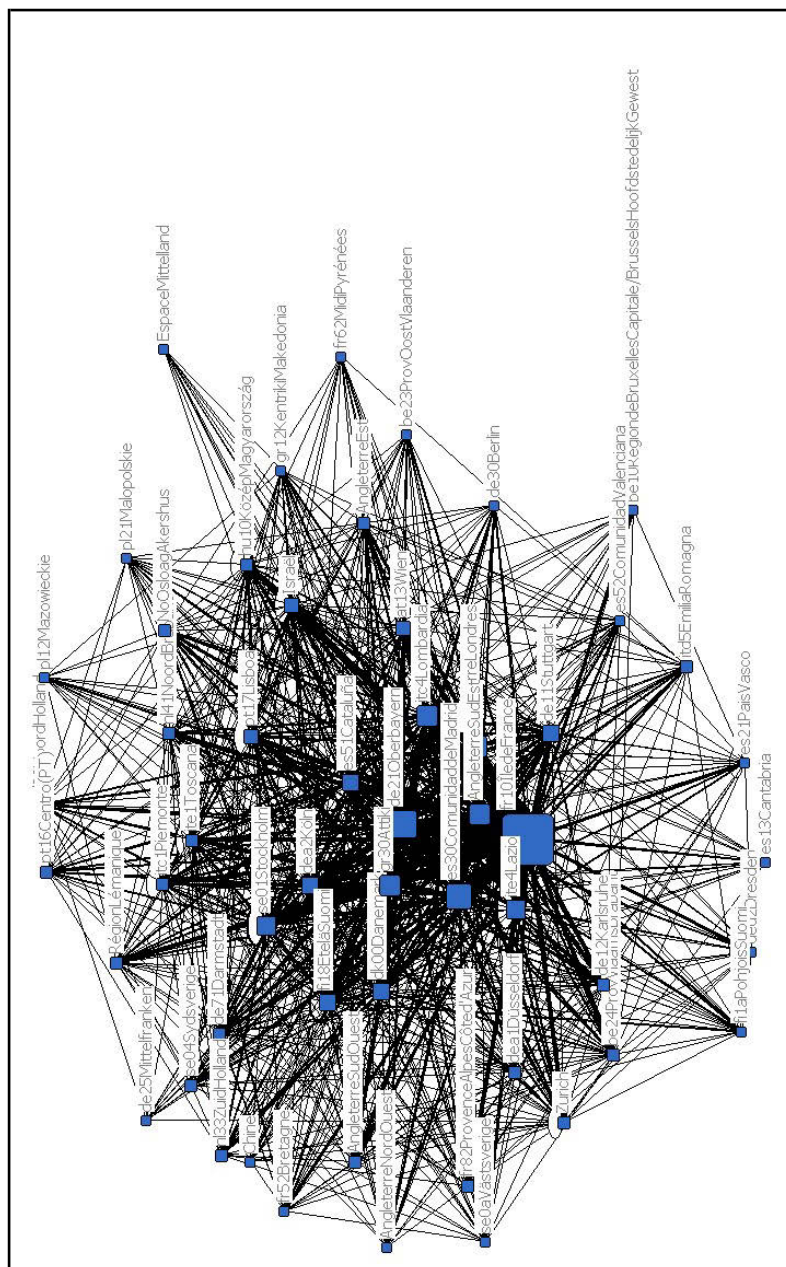
(*UE15 = Union Européenne à 15 membres; **NEM = Nouveaux États Membres; ***UE27 = Union Européenne à 27 membres)

Tableau : Indicateurs de participations de chaque pays au réseau dans le réseau Télécommunications.

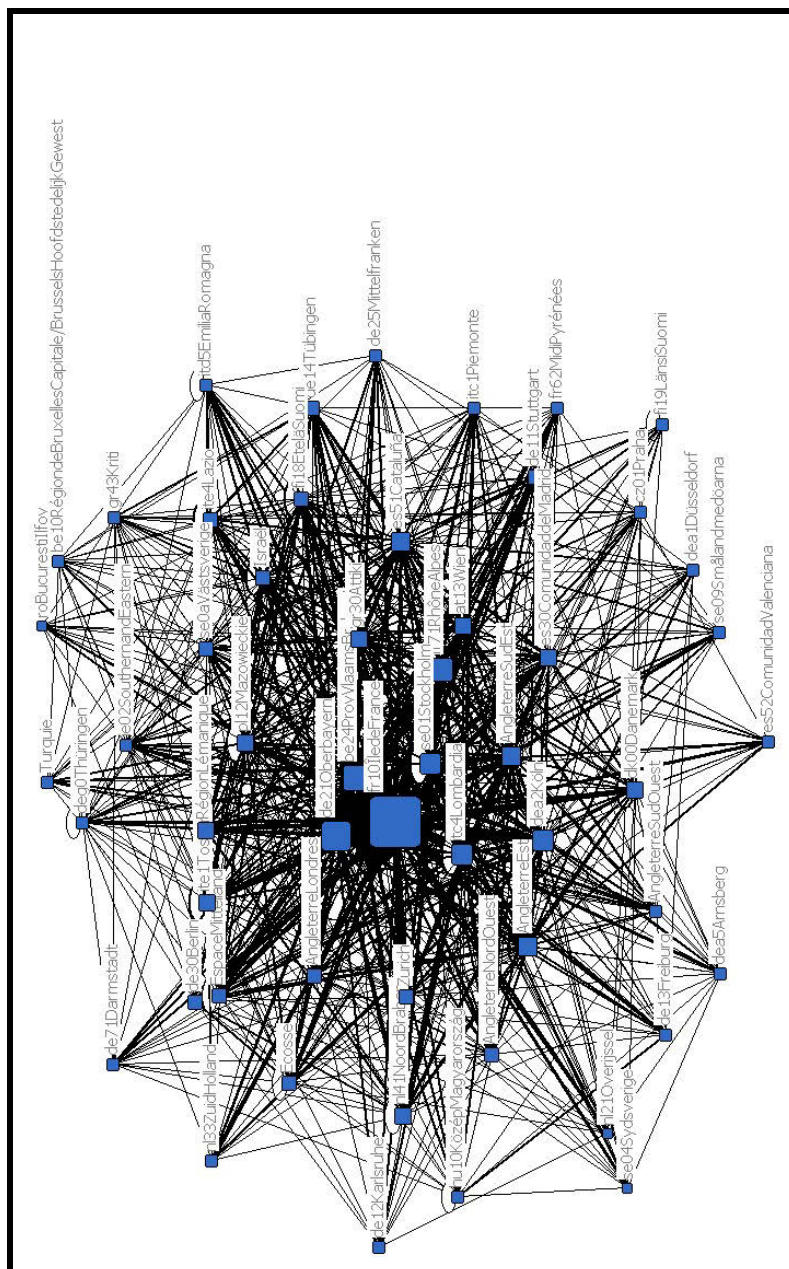
Pays	Nombre d'agents	Répartition en %	Nombre de projets	Répartition en %	Participation Moyenne
Ukraine	0	0,00	0	0,00	-
Islande	1	0,08	1	0,03	1,00
Norvège	20	1,65	52	1,75	2,60
Suisse	32	2,64	92	3,09	2,88
Argentine	1	0,08	1	0,03	1,00
Australie	2	0,17	3	0,10	1,50
Biélorussie	0	0,00	0	0,00	-
Brésil	4	0,33	4	0,13	1,00
Canada	4	0,33	10	0,34	2,50
Chine	9	0,74	13	0,44	1,44
Corée du Sud	4	0,33	5	0,17	1,25
Croatie	1	0,08	3	0,10	3,00
Etats-Unis	7	0,58	7	0,23	1,00
Inde	1	0,08	3	0,10	3,00
Israël	29	2,39	42	1,41	1,45
Japon	7	0,58	7	0,23	1,00
Liechtenstein	0	0,00	0	0,00	-
Russie	6	0,50	7	0,23	1,17
Serbie	1	0,08	1	0,03	1,00
Singapour	3	0,25	10	0,34	3,33
Taïwan	1	0,08	1	0,03	1,00
Turquie	8	0,66	11	0,37	1,38
Total/Moyenne	1211	100	2979	100	2,46

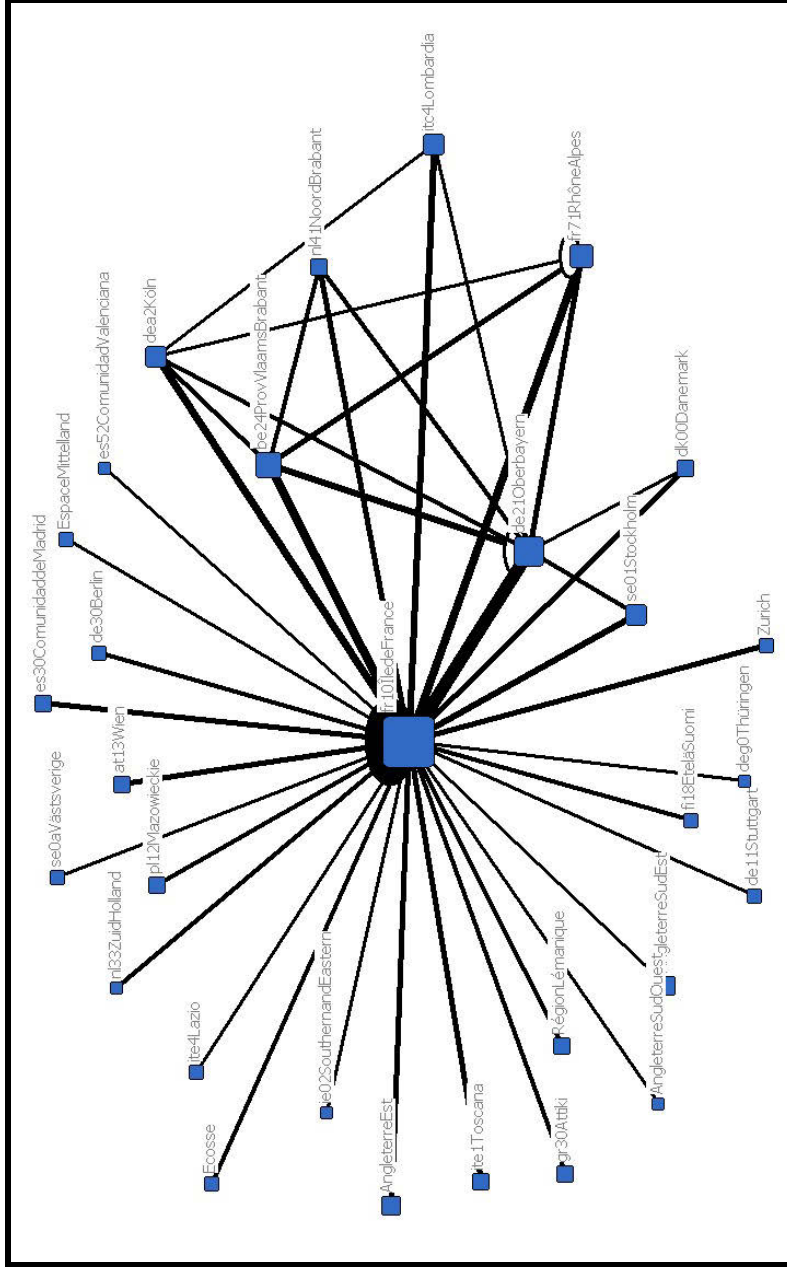
Tableau bis : Indicateurs de participations de chaque pays au réseau dans le réseau Télécommunications.

Annexes A.10.1 : Cœur des réseaux de régions Télécommunications et Electronique-Microélectronique



Graphes A : Cœur du réseau Télécommunications. La taille des points varie en fonction du degré des agents et la taille des liens en fonction de leur intensité.





Graphe D : Cœur du réseau du réseau Electronique-Microélectronique. La taille des points varient en fonction du degré des agents et la taille des liens en fonction de leur intensité.

Annexes A.11.1 : Cœur du réseaux Télécommunications par la méthode cœur/périphérie

Régions	Régions
si00 Slovénie	ded2 Dresden
se0a Västsverige	dea2 Köln
se09 Smålandmedöarna	dea1 Düsseldorf
se04 Sydsverige	de71 Darmstadt
se01 Stockholm	de50 Bremen
ro BucurestiIlfov	de30 Berlin
pt17 Lisboa	de25 Mittelfranken
pt16 Centro(PT)	de21 Oberbayern
pt11 Norte	de12 Karlsruhe
pl21 Malopolskie	de11 Stuttgart
pl12 Mazowieckie	be31 ProvBrabantWallon
nl41 NoordBrabant	be24 ProvVlaamsBrabant
nl33 ZuidHolland	be23 ProvOostVlaanderen
nl32 NoordHolland	be21 ProvAntwerpen
nl21 Overijssel	be10 RégiondeBruxellesCapitale
ite4 Lazio	at31 Oberösterreich
ite1 Toscana	at22 Steiermark
itd5 EmiliaRomagna	at13 Wien
itc4 Lombardia	Zurich
itc1 Piemonte	Turquie
hu10 KözépMagyarország	Singapour
gr30 Attiki	Région Lémanique
gr12 KentrikiMakedonia	No Trondelag
fr82 ProvenceAlpesCôte d'Azur	No Osloag Akershus
fr62 MidiPyrénées	Israël
fr52 Bretagne	Espace Mittelland
fr10 Île de France	Ecosse
fi1a Pohjois Suomi	Chine
fi18 Etelä Suomi	Canada
es52 Comunidad Valenciana	Angleterre Yorkshire et Humber
es51 Cataluña	Angleterre Sud Ouest
es30 Comunidad de Madrid	Angleterre Sud Est
es21 Pais Vasco	Angleterre Nord Ouest
es13 Cantabria	Angleterre Londres
dk00 Danemark	Angleterre Est

**Annexes A.11.2 : Cœur du réseaux Electronique-Microélectronique
par la méthode cœur/périphérie**

Régions
AngleterreEst
Angleterre Sud Est
Angleterre Sud Ouest
Ecosse
Espace Mittelland
Israël
Région Lémanique
Zurich
at13 Wien
be24 ProvVlaamsBrabant
de11 Stuttgart
de21 Oberbayern
dea 2Köln
dk00 Danemark
es30 Comunidad de Madrid
es51 Cataluña
fi18 Etelä Suomi
fr10 Île de France
fr71 Rhône Alpes
gr30 Attiki
itc4 Lombardia
ite1 Toscana
nl41 NoordBrabant
pl12 Mazowieckie
se01 Stockholm

Annexes A.12 : Performance des régions au sein des PCRD

Régions	Degré	Rang	Nombre d'agents	Rang	Nombre de participations	Rang	Av. rank	Participation moyenne	Effectif de R&D en ETP*
Île de France	7162	1	101	1	304	1	1	3,01	135219
Oberbayern	3189	2	28	9	159	2	2,7	5,68	60705,5
Comunidad de Madrid	2824	3	36	3	129	3	3	3,58	41129,2
Angleterre Sud-Est	2327	4	38	2	100	5	3,16	2,63	65079,5
Lombardia	2226	6	36	3	91	7	4,67	2,53	30343
Attiki	2288	5	33	7	104	4	5,06	3,15	18489
Angleterre Londres	2061	7	35	5	98	6	5,89	2,80	41807
Lazio	1700	9	35	5	78	8	6,88	2,23	30419,66667
Cataluña	1426	12	22	11	61	9	10,51	2,77	35361,6
Stockholm	1720	8	16	15	56	12	10,91	3,50	25883
Etelä-Suomi	1557	10	17	14	59	10	11,05	3,47	33758
Köln	1542	11	20	13	59	10	11,20	2,95	30007,5
Danemark	1373	13	23	10	51	13	11,82	2,22	43015,4
Israël	997	16	29	8	42	16	12	1,45	nd
Lisboa	1060	15	15	16	45	15	15,32	3,00	13124
Stuttgart	1303	14	14	19	48	14	15,35	3,43	49336,5
Darmstadt	848	21	21	12	36	18	16,09	1,71	31143,5
Piemonte	812	23	15	16	36	18	18,57	2,40	18474,33333
Közép-Magyarország	857	20	14	19	38	17	18,58	2,71	15202,8
Wien	929	17	13	21	35	20	19,18	2,69	17713,66667

(*Effectif de R&D en ETP : Ensemble du personnel et des chercheurs de R&D tout secteur d'exécution confondus par région mesuré en équivalent temps plein ; nd : données non disponibles ou données insuffisantes pour le calcul de l'indicateur.)

Tableau 1 : Classement des 20 régions les plus centrales du réseau Télécommunications en fonction des indicateurs de participations absolus.

Annexes A.12 : Performance des régions au sein des PCRD

Régions	Degré	Nombre d'agents	Nombre de participations	Participation moyenne	Rang	Effectif de R&D en ETP*	Indicateur de performance	Rang	Average Rank
Oberbayern	3189	28	159	5,68	1	60705,5	0,0525	2	1,33
Île de France	7162	101	304	3,01	20	135219	0,0530	1	1,90
Düsseldorf	641	4	20	5,00	2	20009	0,0320	26	3,71
Angleterre Sud-Est	2327	38	100	2,63	30	65079,5	0,0358	2	3,75
Noord-Brabant	757	7	32	4,57	3	16419	0,0461	34	5,51
Stuttgart	1303	14	48	3,43	13	49336,5	0,0264	4	6,12
Pohjois-Suomi	508	4	18	4,50	4	7424,8	0,0684	59	7,49
Centro (PT)	571	6	27	4,50	4	4675,2	0,1221	70	7,57
Comunidad de Madrid	2824	36	129	3,58	10	41129,2	0,0687	7	8,24
Angleterre Est	638	10	23	2,30	39	45687,5	0,0140	5	8,86
Danemark	1373	23	51	2,22	44	43015,4	0,0319	5	8,98
Karlsruhe	886	8	32	4,00	7	28347,5	0,0313	18	10,08
Angleterre Londres	2061	35	98	2,80	23	41807	0,0493	7	10,73
Etelä-Suomi	1557	17	59	3,47	12	33758	0,0461	10	10,91
Trondelag	268	3	13	4,33	6	4523	0,0593	73	11,09
Turquie	315	8	11	1,38	72	42185,4	0,0075	7	12,76
Sydsverige	626	4	15	3,75	8	12492	0,0501	47	13,67
Cataluña	1426	22	61	2,77	24	35361,6	0,0403	10	14,12
Zuid-Holland	592	6	22	3,67	9	17521	0,0338	33	14,14
Stockholm	1720	16	56	3,50	11	25883	0,0665	20	14,19

(*Effectif de R&D en ETP : Ensemble du personnel et des chercheurs de R&D tout secteur d'exécution confondus par région mesuré en équivalent temps plein ; nd : données non disponibles ou données insuffisantes pour le calcul de l'indicateur.)

Tableau 1 bis : Classement des 20 régions les plus centrales du réseau Télécommunications en fonction des indicateurs de participations relatifs.

Annexes A.12 : Performance des régions au sein des PCRD

Régions	Degré	Rang	Nombre d'agents	Rang	Nombre de participations	Rang	Av. rank	Participation moyenne	Effectif de R&D en ETP*
Île de France	2706	1	65	1	171	1	1	2,63	135219
Oberbayern	1359	2	19	7	95	2	2,63	5,00	60705,5
Rhône-Alpes	889	4	35	2	51	4	3	1,46	38046
Prov. Vlaams Brabant	1043	3	6	50	72	3	4,37	12,00	nd
Angleterre Sud-Est	656	8	27	3	42	7	4,99	1,56	65079,5
Lombardia	767	7	21	6	51	4	5,36	2,43	30343
Stockholm	843	5	15	11	44	6	6,56	2,93	25883
Angleterre Est	641	10	22	5	42	7	6,77	1,91	45687,5
Israël	442	19	24	4	29	14	8,02	1,21	nd
Köln	774	6	14	12	40	9	8,31	2,86	30007,5
Danemark	524	14	19	7	32	12	10,08	1,68	43015,4
Comunidad de Madrid	582	11	12	16	34	11	12,28	2,83	41129,2
Toscana	507	15	16	10	24	21	14	1,50	11014,3
Cataluña	653	9	10	25	24	21	15,10	2,40	35361,6
Attiki	578	12	10	25	29	14	15,40	2,90	18489,0
Ecosse	358	28	18	9	24	21	15,43	1,33	26519,0
Etelä-Suomi	439	20	13	13	27	16	15,84	2,08	33758,0
Noord-Brabant	495	17	7	40	39	10	16,32	5,57	16419,0
Wien	570	13	11	18	23	24	17,23	2,09	17713,7
Région Lémanique	481	18	9	31	32	12	17,53	3,56	nd

(*Effectif de R&D en ETP : Ensemble du personnel et des chercheurs de R&D tout secteur d'exécution confondus par région mesuré en équivalent temps plein ; nd : données non disponibles ou données insuffisantes pour le calcul de l'indicateur.))

Tableau 2 : Classement des 20 régions les plus centrales du réseau Electronique-Microélectronique en fonction des indicateurs de participations absolus.

Annexes A.12 : Performance des régions au sein des PCRDT

Régions	Degré	Nombre d'agents	Nombre de participations	Participation moyenne	Rang	Effectif de R&D en ETP*	Indicateur de performance	Rang	Average Rank
Noord-Brabant	495	7	39	5,57	1	16419	0,030	11	1,83
Småland med öarna	246	6	11	1,83	36	2221	0,111	1	1,95
Kriti	225	4	10	2,50	16	2062,5	0,109	2	3,56
Oberbayern	1359	19	95	5,00	2	60705,5	0,022	22	3,67
Zuid-Holland	260	3	14	4,67	3	17521	0,015	47	5,64
Toscana	507	16	24	1,50	55	11014,33	0,046	3	5,69
Dytiki Ellada	91	3	7	2,33	23	2223	0,041	4	6,81
Comunidad Valenciana	247	4	15	3,75	4	14281,4	0,017	39	7,26
Stockholm	843	15	44	2,93	7	25883	0,033	8	7,47
Overijssel	202	7	17	2,43	19	5005,5	0,040	5	7,92
Mazowieckie	506	8	25	3,13	5	24934,2	0,020	29	8,53
Attiki	578	10	29	2,90	8	18489	0,031	10	8,89
Praha	240	4	12	3,00	6	14416,2	0,017	42	10,50
Saarland	105	5	6	1,20	76	2744	0,038	6	11,12
Köln	774	14	40	2,86	9	30007,5	0,026	15	11,25
Thüringen	332	13	16	1,23	74	9211	0,036	7	12,79
Wien	570	11	23	2,09	30	17713,67	0,032	9	13,85
Västsvrige	400	7	19	2,71	11	17352	0,023	21	14,44
Liguria	121	3	8	2,67	13	4956	0,024	17	14,73
Comunidad de Madrid	582	12	34	2,83	10	41129,2	0,014	49	16,61

(*Effectif de R&D en ETP : Ensemble du personnel et des chercheurs de R&D tout secteur d'exécution confondus par région mesuré en équivalent temps plein ; nd : données non disponibles ou données insuffisantes pour le calcul de l'indicateur.)

Tableau 2 bis : Classement des 20 régions les plus centrales du réseau Electronique-Microélectronique en fonction des indicateurs de participations relatifs.

Annexes A.12 : Performance des régions au sein des PCRD

Régions	Degré	Rang	Nombre d'agents	Rang	Nombre de participations	Rang	Av. rank	Participation moyenne	Effectif de R&D en ETP*
Île de France	9868	1	141	1	475	1	1	3,37	135219
Oberbayern	4548	2	40	8	254	2	2,67	6,35	60705,5
Angleterre Sud-Est	2983	5	59	2	142	4	3,16	2,41	65079,5
Comunidad de Madrid	3406	3	43	5	163	3	3,46	3,79	41129,2
Lombardia	2993	4	48	4	142	4	4	2,96	30343
Israël	1439	17	49	3	71	16	6,60	1,45	nd
Attiki	2866	6	38	10	133	6	6,92	3,50	18489
Angleterre Londres	2422	8	42	7	121	7	7,30	2,88	41807
Stockholm	2563	7	26	16	100	9	9,48	3,85	25883
Lazio	2105	10	39	9	95	11	9,93	2,44	30419,67
Köln	2316	9	31	12	99	10	10,19	3,19	30007,5
Rhône-Alpes	1102	28	43	5	63	19	10,40	1,47	38046
Cataluña	2079	11	28	13	85	13	12,26	3,04	35361,6
Danemark	1897	13	35	11	83	14	12,54	2,37	43015,4
Etelä-Suomi	1996	12	26	16	86	12	13,09	3,31	33758
Prov. Vlaams Brabant	1792	14	9	59	105	8	14,06	11,67	nd
Stuttgart	1660	15	22	19	75	15	16,13	3,41	49336,5
Angleterre Est	1279	20	27	15	65	18	17,42	2,41	45687,5
Darmstadt	1114	26	28	13	52	24	19,10	1,86	31143,5
Wien	1499	16	20	22	58	21	19,28	2,90	17713,67

(*Effectif de R&D en ETP : Ensemble du personnel et des chercheurs de R&D tout secteur d'exécution confondus par région mesuré en équivalent temps plein ; nd : données non disponibles ou données insuffisantes pour le calcul de l'indicateur.)

Tableau 3 : Classement des 20 régions les plus centrales du réseau agréé en fonction des indicateurs de participations absolus.

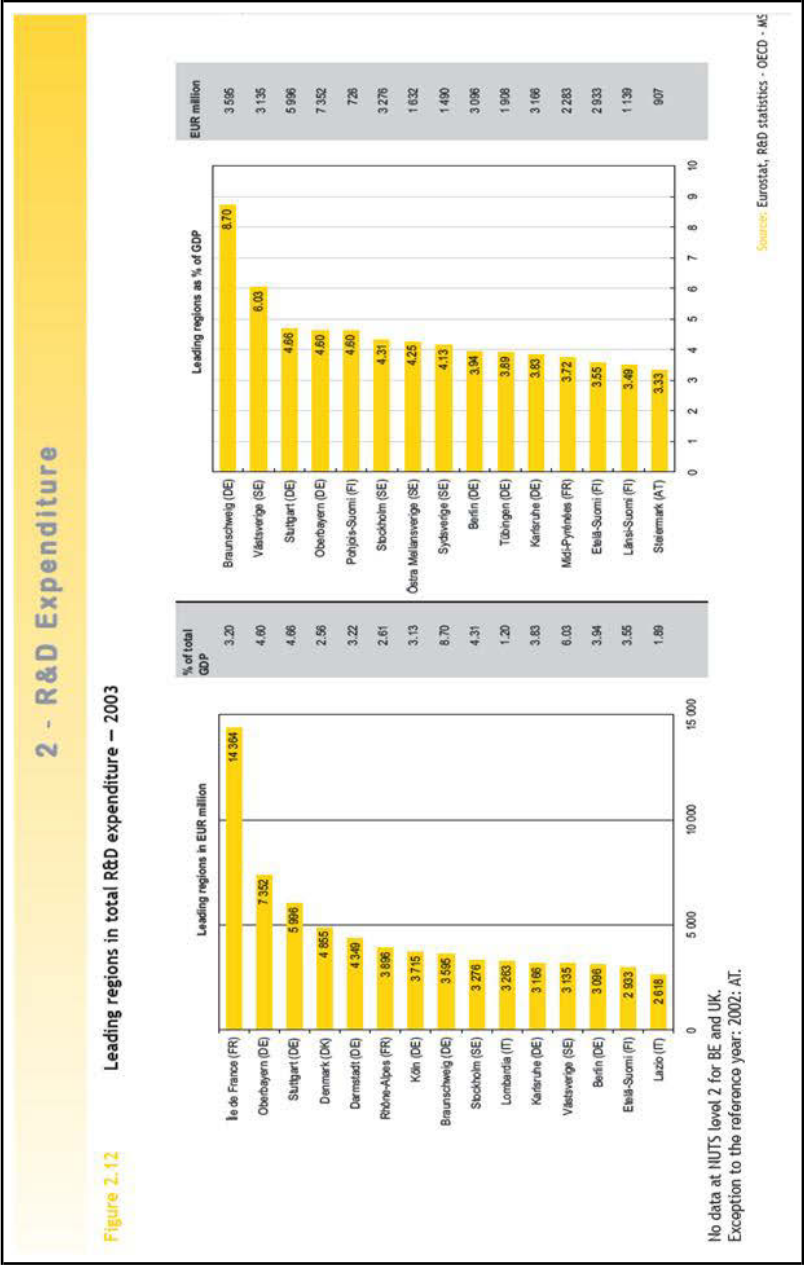
Annexes A.12 : Performance des régions au sein des PCRDT

Régions	Degré	Nombre d'agents	Nombre de participations	Participation moyenne	Rang	Effectif de R&D en ETP*	Indicateur de performance	Rang	Average Rank
Oberbayern	4548	40	254	6,35	1	60705,5	0,075	23	1,917
Småland med öarna	480	13	21	1,62	86	2221	0,216	1	1,977
Attiki	2866	38	133	3,50	14	18489	0,155	2	3,5
Noord-Brabant	1252	12	71	5,92	2	16419	0,076	21	3,652
Pohjois-Suomi	560	4	22	5,50	3	7424,8	0,075	22	5,28
Centro (PT)	695	8	31	3,88	8	4675,2	0,149	4	5,333
Kriti	318	6	18	3,00	27	2062,5	0,154	3	5,4
Comunidad Valenciana	746	6	32	5,33	4	14281,4	0,052	39	7,256
Dytiki Ellada	314	5	17	3,40	17	2223	0,141	5	7,727
Trondelag	373	4	19	4,75	6	4523	0,082	17	8,870
Zuid-Holland	852	7	36	5,14	5	17521	0,049	43	8,958
Overijssel	563	10	34	3,40	17	5005,5	0,112	7	9,917
Stockholm	2563	26	100	3,85	9	25883	0,099	12	10,286
Kentriki Makedonia	622	7	24	3,43	15	5549,5	0,112	8	10,435
Toscana	1338	25	50	2,00	60	11014,33333	0,121	6	10,909
Comunidad de Madrid	3406	43	163	3,79	10	41129,2	0,083	16	12,308
Brandenburg - Südwest	201	3	12	4,00	7	4736,5	0,042	52	12,339
Chypre	115	4	9	2,25	56	1028,8	0,112	9	15,508
Kärnten	213	5	11	2,20	59	1945	0,110	10	17,101
Sydsverige	824	7	25	3,57	12	12492	0,066	30	17,143

(*Effectif de R&D en ETP : Ensemble du personnel et des chercheurs de R&D tout secteur d'exécution confondus par région mesuré en équivalent temps plein ; nd : données non disponibles ou données insuffisantes pour le calcul de l'indicateur.)

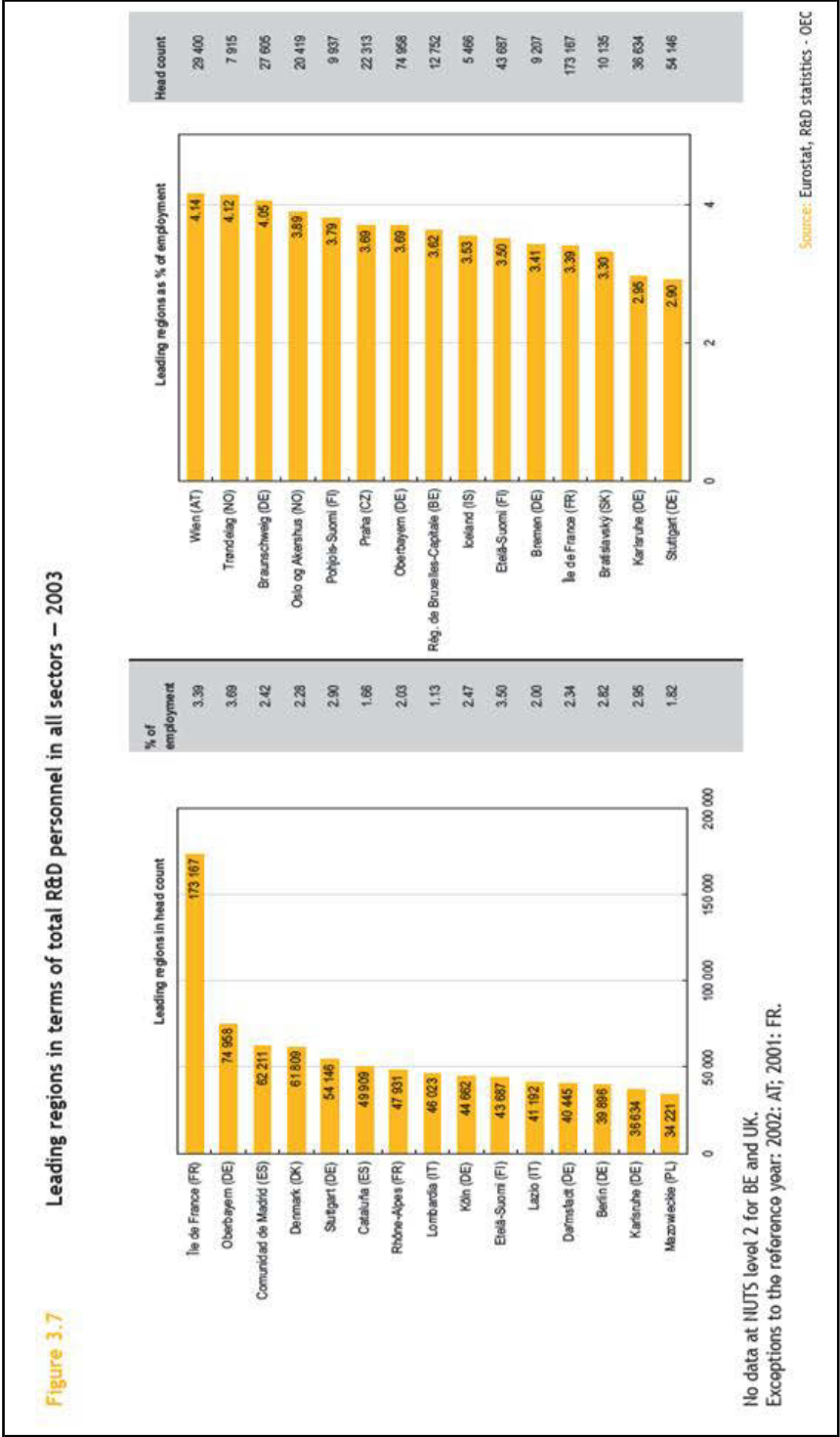
Tableau 3 bis : Classement des 20 régions les plus centrales du réseau agrégé en fonction des indicateurs de participations relatifs.

Annexes A.12 : Performance des régions au sein des PCRD



Graphique 1 : Régions qui dépendent le plus en R&D en Europe.
Source : Eurostat (2007), “Science, technology and innovation in Europe”.

Annexes A.12 : Performance des régions au sein des PCRD



Graphique 2 : Régions qui possèdent le plus de personnel chercheurs en Europe.
Source : Eurostat (2007), “Science, technology and innovation in Europe”.

Annexes A.12 : Performance des régions au sein des PCRDT

Rang	Région	Pays	Score
1	Stockholm	Suède	0,9
2	Västsverige	Suède	0,83
3	Oberbayern	Allemagne	0,79
4	Etelä-Suomi	Suède	0,78
5	Karlsruhe	Allemagne	0,77
6	Stuttgart	Allemagne	0,77
7	Braunschweig	Allemagne	0,76
8	Sydsverige	Suède	0,76
9	Ile de France	France	0,75
10	Ostra Mellansverige	Suède	0,74
11	Berlin	Allemagne	0,74
12	South East	Royaume-Uni	0,72
13	Tübingen	Allemagne	0,72
14	Manner-Suomi	Finlande	0,71
15	Praha	République Tchèque	0,7
16	Darmstadt	Allemagne	0,69
17	Eastern	Royaume-Uni	0,69
18	Dresden	Allemagne	0,69
19	Köln	Allemagne	0,69
20	Noord-Brabant	Pays-Bas	0,68
20	Denmark	Danemark	0,68
20	Pohjols-Suomi	Finlande	0,68
20	Mittelfranken	Allemagne	0,68
20	Wien	Autriche	0,68

Tableau D : Classement des 20 premières régions selon l'indicateur "Regional Performance Innovation 2006" (Source : Inno Metrics, 2006).

Annexes A.12 : Performance des régions au sein des PCRDT

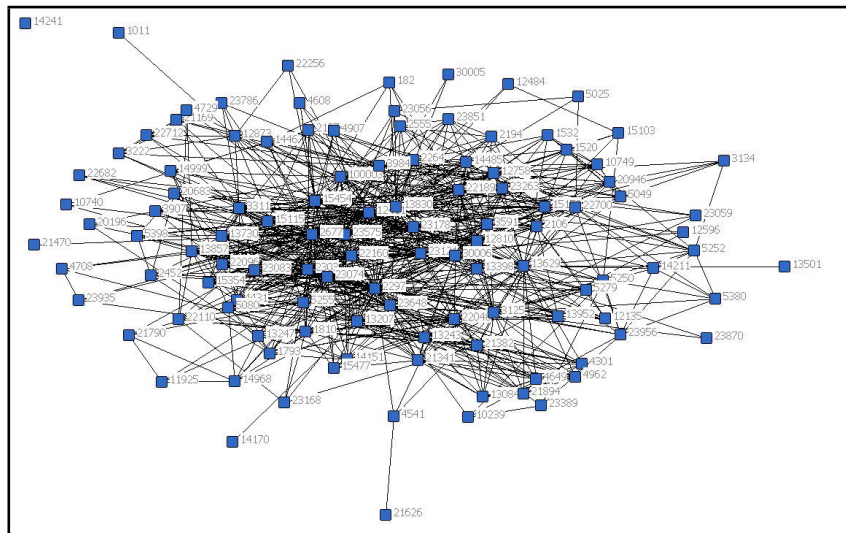
1 Stockholm (SE)	0.90	73 Piemonte (IT)	0.49	141 Poitou-Charentes (FR)	0.34
2 Västsvetige (SE)	0.83	74 Düsseldorf (DE)	0.49	142 Galicia (ES)	0.34
3 Oberbayern (DE)	0.79	75 Provence-Alpes-Côte d'Azur (FR)	0.49	143 Lithuania (LT)	0.33
4 Etelä-Suomi (FI)	0.78	76 Comunidad Foral de Navarra (ES)	0.48	144 Közép-Dunántúl (HU)	0.33
5 Karlsruhe (DE)	0.77	77 Southern and Eastern (IE)	0.48	145 Cyprus (CY)	0.32
6 Stuttgart (DE)	0.77	78 North East (UK)	0.48	146 Champagne-Ardenne (FR)	0.32
7 Braunschweig (DE)	0.76	79 Luxembourg (Grand-Duché) (LU)	0.48	147 Weser-Ems (DE)	0.32
8 Sydsvetige (SE)	0.76	80 Wales (UK)	0.48	148 Latvia (LV)	0.32
9 Ile de France (FR)	0.75	81 Emilia-Romagna (IT)	0.47	149 Malta (MT)	0.31
10 Östra Mellansverige (SE)	0.74	82 Cataluña (ES)	0.47	150 Střední Morava (CZ)	0.31
11 Berlin (DE)	0.74	83 Tirol (AT)	0.47	151 Południowo-Zachodni (PL)	0.31
12 South East (UK)	0.72	84 Brandenburg (DE)	0.47	152 Campania (IT)	0.31
13 Tübingen (DE)	0.72	85 Centre (FR)	0.46	153 Centro (PT) (PT)	0.31
14 Manner-Suomi (FI)	0.71	86 Attiki (GR)	0.46	154 Åland (FI)	0.30
15 Praha (CZ)	0.70	87 Picardie (FR)	0.46	155 Łódzkie (PL)	0.29
16 Darmstadt (DE)	0.69	88 Chemnitz (DE)	0.46	156 Śląskie (PL)	0.29
17 Eastern (UK)	0.69	89 Scotland (UK)	0.45	157 Burgenland (AT)	0.29
18 Dresden (DE)	0.69	90 Aragón (ES)	0.45	158 Región de Murcia (ES)	0.29
19 Köln (DE)	0.69	91 Schleswig-Holstein (DE)	0.45	159 Basilicata (IT)	0.29
20 Noord-Brabant (NL)	0.68	92 Oberösterreich (AT)	0.45	160 Dessau (DE)	0.29
21 Denmark (DK)	0.68	93 Languedoc-Roussillon (FR)	0.44	161 Lubelskie (PL)	0.27
22 Pohjois-Suomi (FI)	0.68	94 Liguria (IT)	0.44	162 Północny (PL)	0.27
23 Mittelfranken (DE)	0.68	95 Friuli-Venezia Giulia (IT)	0.44	163 Cantabria (ES)	0.27
24 Wien (AT)	0.68	96 Saarland (DE)	0.44	164 Kentriki Makedonia (GR)	0.27
25 Utrecht (NL)	0.66	97 Oberfranken (DE)	0.44	165 Molise (IT)	0.27
26 Rheinhessen-Pfalz (DE)	0.66	98 Aquitaine (FR)	0.44	166 Principado de Asturias (ES)	0.27
27 Bratislavský kraj (SK)	0.66	99 Vorarlberg (AT)	0.43	167 Stredné Slovensko (SK)	0.27
28 Länsi-Suomi (FI)	0.65	100 Jihovýchod (CZ)	0.43	168 Corse (FR)	0.26
29 Freiburg (DE)	0.63	101 Stredni Cecly (CZ)	0.43	169 Andalusia (ES)	0.26
30 Midi-Pyrénées (FR)	0.61	102 Kärnten (AT)	0.43	170 Valle d'Aosta/Vallee d'Aoste (IT)	0.26
31 Comunidad de Madrid (ES)	0.61	103 Amsberg (DE)	0.43	171 Západné Slovensko (SK)	0.26
32 Vlaams Gewest (BE)	0.61	104 Toscana (IT)	0.43	172 Północno-Zachodni (PL)	0.26
33 Rhône-Alpes (FR)	0.60	105 Detmold (DE)	0.43	173 Észak-Alföld (HU)	0.26
34 Közép-Magyarország (HU)	0.60	106 Pays de la Loire (FR)	0.42	174 Kriš (GR)	0.26
35 London (UK)	0.59	107 Umbria (IT)	0.42	175 Dél-Dunántúl (HU)	0.26
36 Flevoland (NL)	0.59	108 Lisboa (PT)	0.42	176 Nyugat-Dunántúl (HU)	0.25
37 South West (UK)	0.58	109 Abruzzo (IT)	0.42	177 Sicilia (IT)	0.25
38 Zuid-Holland (NL)	0.58	110 Halle (DE)	0.42	178 Észak-Magyarország (HU)	0.25
39 Gelderland (NL)	0.58	111 Auvergne (FR)	0.42	179 Dél-Alföld (HU)	0.24
40 Noord-Holland (NL)	0.58	112 Limousin (FR)	0.42	180 Moravskoslezsko (CZ)	0.24
41 Steiermark (AT)	0.58	113 Northern Ireland (UK)	0.41	181 La Rioja (ES)	0.23
42 West Midlands (UK)	0.57	114 Niederbayern (DE)	0.41	182 Dytiki Ellada (GR)	0.23
43 Leipzig (DE)	0.57	115 Trier (DE)	0.41	183 Canarias (ES) (ES)	0.23
44 Lazio (IT)	0.57	116 Salzburg (AT)	0.41	184 Sardegna (IT)	0.23
45 Norra Mellansverige (SE)	0.57	117 Münster (DE)	0.41	185 Puglia (IT)	0.22
46 Övre Norrland (SE)	0.57	118 Haute-Normandie (FR)	0.41	186 Norte (PT)	0.22
47 East Midlands (UK)	0.57	119 Kassel (DE)	0.41	187 Podkarpackie (PL)	0.21
48 Schwaben (DE)	0.56	120 Basse-Normandie (FR)	0.41	188 Calabria (IT)	0.20
49 Gießen (DE)	0.56	121 Lorraine (FR)	0.40	189 Východné Slovensko (SK)	0.19
50 Hannover (DE)	0.56	122 Veneto (IT)	0.40	190 Algarve (PT)	0.19
51 Alsace (FR)	0.55	123 Drenthe (NL)	0.38	191 Ipeiros (GR)	0.19
52 Unterfranken (DE)	0.55	124 Estonia (EE)	0.38	192 Sterea Ellada (GR)	0.17
53 Hamburg (DE)	0.55	125 Koblenz (DE)	0.38	193 Extremadura (ES)	0.17
54 Oberpfalz (DE)	0.55	126 Lüneburg (DE)	0.38	194 Castilla-La Mancha (ES)	0.17
55 Pais Vasco (ES)	0.55	127 Mecklenburg-Vorpommern (DE)	0.37	195 Illes Balears (ES)	0.16
56 North West (UK)	0.54	128 Niederösterreich (AT)	0.37	196 Alentejo (PT)	0.13
57 Småland med Öarna (SE)	0.54	129 Bourgogne (FR)	0.36	197 Anatoliki Makedonia, Thraki (GR)	0.13
58 Limburg (NL)	0.53	130 Comunidad Valenciana (ES)	0.36	198 Severozápad (CZ)	0.12
59 Thüringen (DE)	0.53	131 Zeeland (NL)	0.36	199 Peloponnisos (GR)	0.10
60 Bremen (DE)	0.53	132 Marche (IT)	0.35	200 Thessalia (GR)	0.10
61 Groningen (NL)	0.52	133 Border, Midlands and Western (IE)	0.35	201 Dytiki Makedonia (GR)	0.07
62 Région de Bruxelles-Capitale (BE)	0.52	134 Malopolskie (PL)	0.35	202 Voreio Aigaio (GR)	0.04
63 Slovenia (SI)	0.52	135 Castilla y León (ES)	0.35	203 Notio Aigaio (GR)	0.01
64 Overijssel (NL)	0.52	136 Friesland (NL)	0.35		
65 Mazowieckie (PL)	0.51	137 Magdeburg (DE)	0.35		
66 Bretagne (FR)	0.51	138 Jihozápad (CZ)	0.34		
67 Franche-Comté (FR)	0.51	139 Severovýchod (CZ)	0.34		
68 Mellersta Norrland (SE)	0.50	140 Nord - Pas-de-Calais (FR)	0.34		
69 Région Wallonne (BE)	0.49				
70 Itä-Suomi (FI)	0.49				
71 Lombardia (IT)	0.49				
72 Yorkshire and The Humber (UK)	0.49				

Tableau E : Classement des régions européennes selon l'indicateur "Regional Performance Innovation 2006" (Source : Inno Metrics, 2006).
Source : Inno Metrics (2006), "European Innovation Scoreboard".

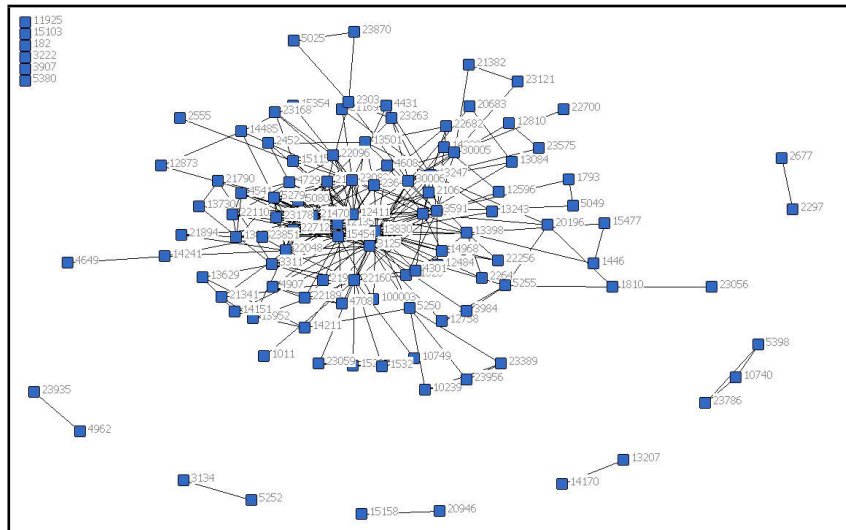
Deuxième partie

Déterminants des choix de collaborations au sein des PCRDT

Annexes A.13. Présentation des hypothèses : Illustration des hypothèses

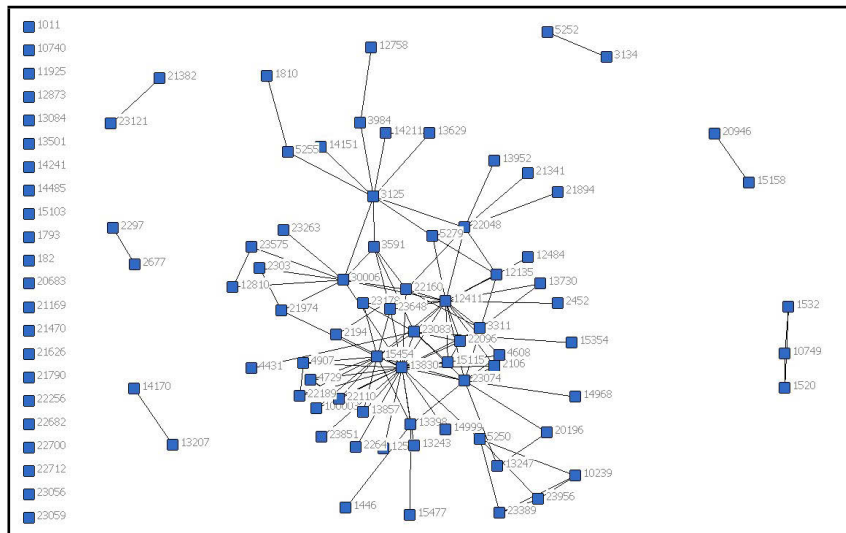


Graphe E : Réseaux des collaborations dans le 6^{ème} PCRDT.

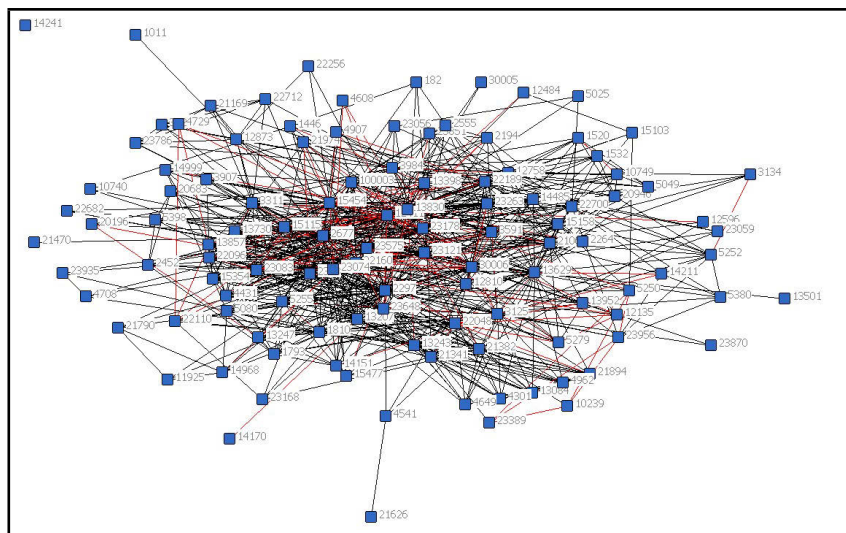


Graphe F : Réseaux des collaborations dans le 5^{ème} PCRDT.

Annexes A.13. Présentation des hypothèses : Illustration des hypothèses

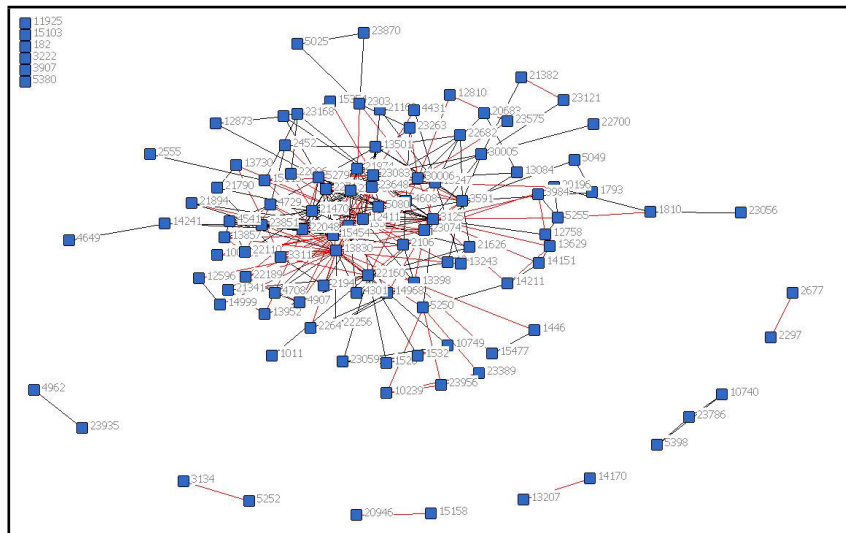


Graphe G : Réseaux des collaborations présentes à la fois dans les 5^{ème} PCRDT et 6^{ème} PCRDT.

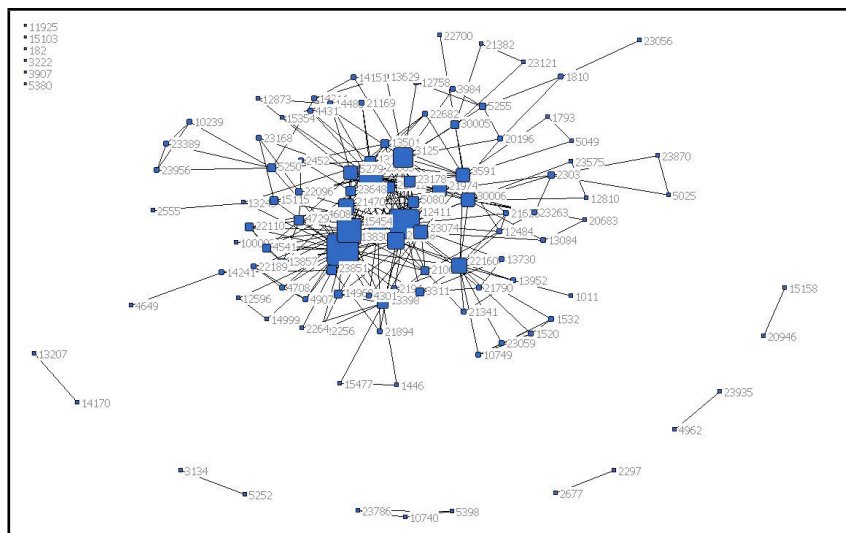


Graphes H : Réseaux des collaborations présentes à la fois dans les 5^{ème} PCRDT et 6^{ème} PCRDT au sein du réseau du 6^{ème} PCRDT.

Annexes A.13. Présentation des hypothèses : Illustration des hypothèses

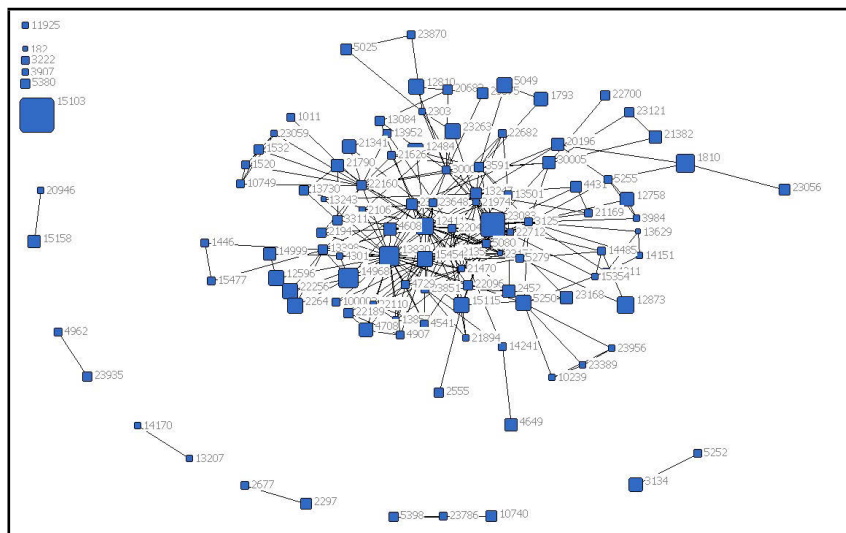


Graphe I : Réseaux des collaborations présentes à la fois dans les 5^{ème} PCRDT et 6^{ème} PCRDT au sein du réseau du 5^{ème} PCRDT.

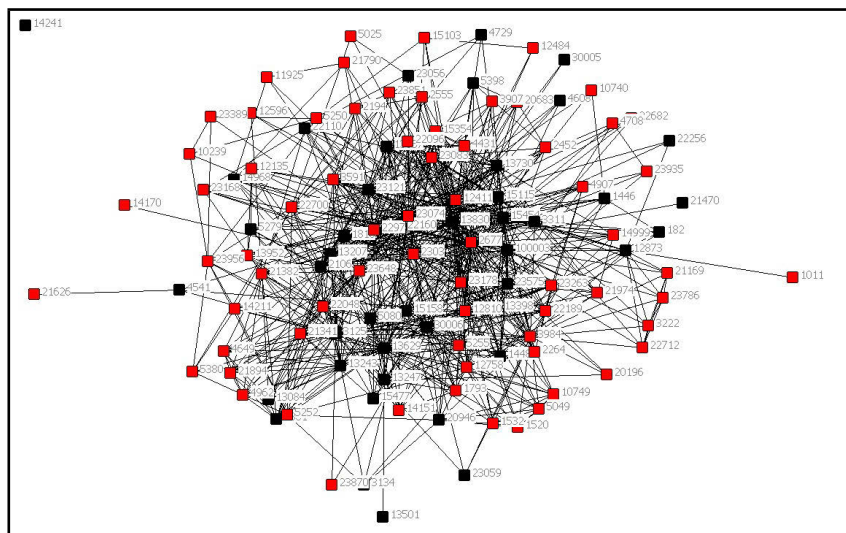


Graphe J : Réseaux des collaborations du 6^{ème} PCRDT. La taille des points varie en fonction de la centralité des agents.

Annexes A.13. Présentation des hypothèses : Illustration des hypothèses



Graphe K : Réseaux des collaborations du 6^{ème} PCRDT. La taille des points varient en fonction du coût moyen par projet supporté par un agent.



Graphe L : Réseaux des collaborations du 6^{ème} PCRDT. La couleur des points varient en fonction de l'appartenance des agents au cœur. En rouge, les agents localisés dans le cœur.

Annexes A.13. Présentation des hypothèses : Résumé des variables et effets attendus

Hypothèses	Variables	Description	Effet escompté
Variable explicative	$LIEN6$	Variable binaire indiquant la présence d'un lien entre les deux agents dans le réseau formé par le 6 ^{ème} PCRDT	
Effets réseau	$LIEN5$	Variable binaire indiquant la présence d'un lien entre les deux agents dans le réseau formé par le 5 ^{ème} PCRDT	+
	$PARTCOM$	Variable indiquant le nombre de partenaires différents communs aux deux agents dans le 5 ^{ème} PCRDT	+
	$DG5$	Variable égale à l'inverse de la distance géodésique entre les deux agents	+
	\overline{DR}	Moyenne du degré renforcé des deux agents dans le réseau complet	+
	ΔDR	Valeur absolue de la différence entre le degré renforcé des deux agents dans le réseau complet	-
Moyens en R&D	\overline{COUT}	Moyenne des coûts supportés par les agents (exprimés en millions d'euros)	+
	$\Delta COUT$	Valeur absolue des différences des coûts supportés par les agents (exprimés en millions d'euros)	-

Annexes A.13 bis. Présentation des hypothèses : Résumé des variables et effets attendus

Hypothèses	Variables	Description	Effet escompté
Effets spatiaux	<i>PROXGEO</i>	Variable binaire indiquant si les deux agents sont issus du même pays ou de pays voisins	+
	<i>DISTANCE</i>	Variable indiquant la distance exprimé en kilomètre qui sépare les deux agents	-
	<i>TEMPS</i>	Variable indiquant le temps exprimé en minute qui sépare les deux agents	-
Espace économique	<i>Core – UE</i>	Variable binaire indiquant si les deux agents appartiennent au coeur de l'UE	+
	<i>Core1/2</i>	Variable binaire indiquant si au moins un des deux agents appartient au coeur de l'UE	+
	<i>UE15</i>	Variable indiquant si les deux agents appartiennent à l'UE composé de 15 membres	+
Type des agents	<i>HERES</i>	Variable binaire indiquant si les deux agents sont des centres de recherche et/ou des agents de l'enseignement supérieur	+
	<i>HE</i>	Variable binaire indiquant si les deux agents sont tous les deux des agents de l'enseignement supérieur	+
	<i>RES</i>	Variable binaire indiquant si les deux agents sont tous les deux des centres de recherche	+
	<i>IND</i>	Variable binaire indiquant si les deux agents sont tous les deux des entreprises	+

Annexes A.14 : Les variables utilisées dans le modèle économétrique : Statistiques descriptives

Variable	Nb*	Moyenne	Écart-type	Somme	Minimum	Maximum
<i>Lien6</i>	6441	0,13352	0,34016	860	0	1
<i>Lien5</i>	6441	0,04751	0,21274	306	0	1
\overline{DR}	6441	267,47423	306,49556	1722801	0	1372
ΔDR	6441	919,53797	201,61207	5922744	1,5	1359
\overline{COST}	6441	0,353	0,357	2273,297	0,000196	2,466
$\Delta COST$	6441	0,443	0,249	2854,687	0,034	2,028
<i>Partcom</i>	6441	0,56032	1,51108	3609	0	18
<i>DG5</i>	6441	0,3511	0,21152	2261	0	1
<i>Temps</i>	6441	755,50815	516,62755	4866228	0	3071
<i>Distance</i>	6441	1179	766,54456	7592669	0	4262
<i>core – UE</i>	6441	0,37494	0,48415	2415	0	1
<i>PROXGEO</i>	6441	0,37463	0,48407	2413	0	1
<i>UE15</i>	6441	0,79972	0,40024	5151	0	1
<i>BORDER</i>	6441	0,27294	0,4455	1758	0	1
<i>IDPAYS</i>						
<i>HE</i>	6441	0,12731	0,33335	820	0	1
<i>HERES</i>	6441	0,30321	0,45968	1953	0	1
<i>IND</i>	6441	0,16069	0,36727	1035	0	1
<i>RES</i>	6441	0,03586	0,18597	231	0	1

Nb = Nombre d'observations

Annexes A.14 : Les variables utilisées dans le modèle économétrique : Coefficient de corrélation de Pearson

ΔDR	ΔDR	DR	$dg5$	$\Delta Cost$	$cost$	he	$heres$	ind	res	$Lien5$	$Lien6$	$Part*$	$Tps*$	$Dist*$	$core*$	$Prox*$	$ue15$	$Bord*$
ΔDR	1.00000	-0.68533	-0.59538	0.01789	-0.08674	-0.03097	-0.00971	0.03475	0.02743	-0.08622	0.11251	-0.15222	-0.03380	-0.02408	0.01969	0.08998	0.03514	0.08333
\bar{DR}	-0.68533	1.00000	<0.001	0.1511	<0.001	0.0129	0.4360	0.0053	0.0277	<0.001	<0.001	<0.001	0.0067	0.0533	0.1141	<0.001	0.0048	<0.001
$dg5$	<0.001	<0.001	1.00000	0.06443	0.17804	0.05736	0.07783	-0.06780	0.01118	0.21309	0.02633	0.39192	0.02667	0.01349	-0.11211	-0.08142	-0.07229	-0.05309
$\Delta cost$	-0.59538	0.78013	<0.001	<0.001	<0.001	<0.001	0.07999	<0.001	0.3695	<0.001	0.0346	<0.001	0.0314	0.2791	<0.001	<0.001	<0.001	<0.001
\overline{cost}	<0.001	<0.001	0.05240	0.13829	0.04768	0.04768	0.07999	-0.03958	0.02176	0.68519	0.10493	0.62398	0.01219	-0.00104	-0.09030	-0.04973	-0.05124	-0.03931
he	<0.001	<0.001	<0.001	<0.001	<0.001	0.0001	<0.001	0.0015	0.0807	<0.001	<0.001	<0.001	0.3279	0.9333	<0.001	<0.001	<0.001	0.0016
$heres$	0.0129	0.06443	0.05240	1.00000	0.69376	-0.17837	-0.18710	0.09745	-0.00577	0.04069	0.00331	0.05133	-0.00910	-0.00758	0.02190	0.04177	0.07241	0.02739
\overline{ind}	0.4360	<0.001	<0.001	0.69376	1.00000	-0.24272	-0.26063	0.26400	0.01020	0.06391	0.03182	0.06270	0.03301	0.5430	0.0788	0.0008	<0.001	0.0280
ind	0.03475	-0.06780	-0.03958	0.09745	0.26400	-0.16712	-0.28864	1.00000	<0.001	0.06391	0.0106	<0.001	0.0801	-0.04044	0.03053	0.05863	0.11976	0.04943
res	0.02743	<0.001	0.0015	<0.001	<0.001	<0.001	<0.001	<0.001	0.4130	<0.001	0.0106	<0.001	0.0801	0.0012	0.0143	<0.001	0.12495	0.10393
$Lien5$	0.0277	0.3695	0.0807	0.6432	0.4130	<0.001	<0.001	<0.001	<0.001	0.3756	0.0012	<0.001	<0.001	<0.001	-0.05239	-0.11086	-0.07073	-0.09809
$Lien6$	-0.08622	0.21309	0.68519	0.04069	0.06391	0.01104	0.04957	-0.00630	0.02758	0.04957	0.14125	0.09092	0.16594	-0.02816	-0.07135	-0.10932	-0.15517	<0.001
$Part$	<0.001	<0.001	<0.001	0.0011	<0.001	0.3756	<0.001	0.6131	0.0269	<0.001	<0.001	0.8050	<0.001	0.0238	0.0012	<0.001	0.3258	0.4712
$Tps*$	0.0067	0.0314	0.3279	0.4651	0.0081	<0.001	<0.001	<0.001	0.07894	1.00000	1.00000	0.14686	0.00252	-0.00353	-0.06171	0.03189	-0.07044	0.02794
$Dist*$	-0.02408	0.01349	-0.00104	-0.00758	-0.04044	0.08663	0.16291	-0.14111	<0.001	0.18055	1.00000	0.14686	0.00252	0.7770	<0.001	0.0105	<0.001	0.0249
$core*$	0.0533	0.2791	0.9333	0.5430	0.0012	<0.001	<0.001	<0.001	<0.001	0.63883	0.14686	1.00000	0.01224	0.00256	-0.01954	-0.01564	-0.01802	-0.00278
$Prox$	-0.01969	-0.11211	-0.09030	0.02190	0.03053	-0.05239	-0.07135	0.01566	0.00308	<0.001	-0.06171	-0.07963	<0.001	-0.32320	1.00000	0.27713	0.38759	0.05893
$ue15$	0.1141	<0.001	<0.001	0.0788	0.0143	<0.001	<0.001	0.2088	0.2333	0.0012	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$Bord*$	0.08998	-0.08142	-0.04973	0.04177	0.05863	-0.11086	-0.10932	0.11552	0.01114	-0.02060	0.00252	0.01224	1.00000	0.94969	-0.35425	-0.61192	0.12445	0.79161
	<0.001	<0.001	<0.001	0.0008	<0.001	<0.001	<0.001	<0.001	0.3712	0.0630	0.0105	0.1169	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	0.03514	-0.07229	-0.05124	0.07241	0.11976	-0.07073	-0.15517	0.12495	<0.001	0.0630	-0.07044	-0.01802	-0.09082	-0.06871	0.38759	1.00000	1.00000	0.02882
	0.0048	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.3258	<0.001	0.1481	<0.001	<0.001	<0.001	<0.001	0.0207	1.00000
	0.08333	-0.05309	-0.03931	0.02739	0.04943	-0.09809	-0.08572	0.10393	0.01490	0.00898	0.02794	-0.00278	-0.41150	-0.38788	0.05893	0.79161	0.02882	1.00000
	<0.001	<0.001	0.0016	0.0280	<0.001	<0.001	<0.001	<0.001	0.2318	0.4712	0.0249	0.8237	<0.001	<0.001	<0.001	<0.001	0.0207	

* : $Dist = DISTANCE$; $Tps = TEMPS$; $core = core - UE$; $Prox = PROXGEO$; $Bord = BORDER$

Annexes A.14 : Les déterminants des collaborations : Estimations

	Modèle 1	Modèle 2	Modèle 3	Modèle 4
<i>CONSTANTE</i>	-5,703*** (0,31)	-4,5712*** (0,3096)	-4,882*** (0,321)	-4,303*** (0,329)
\overline{DR}	0,003*** (0,0002)	0,0009*** (0,0003)	0,0023*** (0,0003)	0,0015*** (0,0003)
ΔDR	0,003*** (0,0002)	0,0026*** (0,0002)	0,0023*** (0,0002)	0,0021*** (0,0002)
\overline{COUT}	0,527*** (0,200)	à,561*** (0,204)	0,622*** (0,202)	0,721*** (0,203)
$\Delta COUT$	-0,452*** (0,144)	-0,484*** (0,147)	-0,512*** (0,147)	-0,536*** (0,147)
<i>DG5</i>		2,662*** (0,228)		
<i>PARTCOM</i>			0,160*** (0,021)	0,474*** (0,054)
<i>PARTCOM</i> ²				-0,028*** (0,005)
<i>PSEUDO – R</i> ²	0,037	0,070	0,057	0,065
<i>R</i> _{VZ} ²	0,064	0,118	0,097	0,110
Nombre d’observations	6441	6441	6441	6441
<i>LIEN6 = 1</i>	860	860	860	860
<i>LIEN6 = 0</i>	5581	5581	5581	5581

Tableau : Estimations des déterminants des choix de collaborations obtenues à partir d’un modèle Logit (6641 observations). Niveau de significativité : *0.10; ** 0.05; *** 0.01 ; ns = non significatif. Les écarts-types sont donnés entre parenthèses. Les valeurs des *PSEUDO – R*² sont arrondies à 10⁻³.

Annexes A.14 : Les déterminants des collaborations : Estimations

	Modèle 5	Modèle 6	Modèle 7	Modèle 8	Modèle 9
<i>CONSTANTE</i>	-4,913*** (0,33)	-4,940*** (0,323)	-4,922*** (0,321)	-4,883*** (0,321)	-4,919*** (0,322)
\overline{DR}	0,0023*** (0,0003)	0,0023*** (0,0003)	0,0023*** (0,0003)	0,0023*** (0,0003)	0,0023*** (0,0003)
ΔDR	0,0023*** (0,0002)	0,0023*** (0,0002)	0,0023*** (0,0002)	0,0023*** (0,0002)	0,0023*** (0,0002)
\overline{COUT}	0,627*** (0,202)	0,632*** (0,202)	0,594*** (0,203)	0,608*** (0,203)	0,617*** (0,202)
$\Delta COUT$	-0,514*** (0,147)	-0,514*** (0,147)	-0,510*** (0,147)	-0,510*** (0,147)	-0,515*** (0,147)
<i>PARTCOM</i>	0,1603*** (0,0212)	0,16*** (0,0212)	0,161*** (0,0212)	0,161*** (0,0212)	0,161*** (0,0212)
<i>DISTANCE</i>	ns				
<i>TEMPS</i>		ns			
<i>PROXGEO</i>			0,132* (0,776)		
<i>BORDER</i>				ns	
<i>IDPAYS</i>					ns
$PSEUDO - R^2$	0,057	0,057	0,057	0,057	
R^2_{VZ}	0,097	0,097	0,097	0,097	
Nombre d'observations	6441	6441	6441	6441	6441
<i>LIEN6</i> = 1	860	860	860	860	860
<i>LIEN6</i> = 0	5581	5581	5581	5581	5581

Tableau : Estimations des déterminants des choix de collaborations obtenues à partir d'un modèle Logit (6641 observations). Niveau de significativité : *0.10 ; ** 0.05 ; *** 0.01 ; ns = non significatif. Les écarts-types sont donnés entre parenthèses. Les valeurs des paramètres ainsi que celles des écarts-types associés sont arrondis à 10^{-2} et les $PSEUDO - R^2$ sont arrondies à 10^{-3} .

Annexes A.14 : Les déterminants des collaborations : Estimations

	Modèle 10	Modèle 11	Modèle 12
<i>CONSTANTE</i>	-4,645*** (0,323)	-4,913 (0,322)	-4,664*** (0,3261)
\overline{DR}	0,0021*** (0,0003)	0,0023*** (0,0003)	0,002*** (0,0003)
ΔDR	0,0023*** (0,0002)	0,0022*** (0,0002)	0,0022*** (0,0002)
\overline{COUT}	0,658*** (0,202)	0,625*** (0,202)	0,616*** (0,204)
$\Delta COUT$	-0,509*** (0,147)	-0,510*** (0,147)	-0,504*** (0,147)
<i>PARTCOM</i>	0,16*** (0,0212)	0,16*** (0,021)	0,161*** (0,021)
<i>Core – UE</i>	-0,3049*** (0,0823)		-0,37*** (0,086)
<i>Core1/2</i>		ns	
<i>PROXGEO</i>			0,23*** (0,081)
<i>PSEUDO – R²</i>	0,059		0,109
R^2_{VZ}	0,101		0,104
Nombre d’observations	6441	6441	6441
<i>LIEN6 = 1</i>	860	860	860
<i>LIEN6 = 0</i>	5581	5581	5581

Tableau : Estimations des déterminants des choix de collaborations obtenues à partir d’un modèle Logit (6641 observations). Niveau de significativité : *0.10 ; ** 0.05 ; *** 0.01 ; ns = non significatif. Les écarts-types sont donnés entre parenthèses. Les valeurs des *PSEUDO – R²* sont arrondies à 10^{-3} .

Annexes A.14 : Les déterminants des collaborations : Statistiques descriptives

Distance géodésique	<i>LIEN</i> 6 = 0	<i>LIEN</i> 6 = 1	Total
Non liés	696	178	874
(en %)	79,6	20,4	100
De 4 à 7	861	59	920
(en %)	93,6	6,4	100
3	2380	225	2605
(en %)	91,4	8,6	100
2	1463	273	1736
(en %)	84,3	15,7	100
1	181	125	306
(en %)	59,2	40,8	100
Total	5581	860	6441

Tableau F : Liens entre la distance géodésique et la variable *LIEN*6.

Distance géodésique	<i>LIEN</i> 6 = 0	<i>LIEN</i> 6 = 1	Total
3 et +	3937	462	4399
(en %)	89,5	10,5	100
2	1463	273	1736
(en %)	84,3	15,7	100
1	181	125	306
(en %)	59,2	40,8	100
Total	5581	860	6441

Tableau G : Liens entre la distance géodésique et la variable *LIEN*6.

Annexes A.14 : Les déterminants des collaborations : Statistiques descriptives

	<i>LIEN6 = 0</i>	<i>LIEN6 = 1</i>	Total
<i>PARTCOM = 0</i> (en %)	4225 89,4	499 10,6	4724 100
<i>PARTCOM > 1</i> (en %)	1356 79,0	361 21,0	1717 100
Total	5581	860	6441

Tableau H : Réalisation des liens selon la présence ou non d'un partenaire commun à la période précédente.

<i>PARTCOM</i>	<i>LIEN6 = 0</i>	<i>LIEN6 = 1</i>	Total
0 (en %)	4225 89,4	499 10,6	4724 100,0
1 (en %)	840 83,9	161 16,1	1001 100
2 (en %)	293 77,9	83 22,1	376 100
3 (en %)	105 72,9	39 27,1	144 100
4 (en %)	36 56,3	28 43,8	64 100
5 (en %)	23 53,5	20 46,5	43 100
6 (en %)	3 27,3	8 72,7	11 100
7 (en %)	2 50	2 50	4 100
8 (en %)	3 75	1 25	4 100
10 (en %)	0	2 100	2 100
11 et + (en %)	51 75	17 25	68 100
Total	5581	860	6441

Tableau I : Réalisation des liens selon le nombre de partenaires communs à la période précédente.

Annexes A.14 : Les déterminants des collaborations : Statistiques descriptives

	<i>LIEN6 = 0</i>	<i>LIEN6 = 1</i>	Total
<i>PROXGEO = 0</i> (en %)	3524 87,5	504 12,5	4028 100
<i>PROXGEO = 1</i> (en %)	2057 85,2	356 14,8	2413 100

Tableau I : Réalisation des liens en fonction de la proximité des agents.

	<i>LIEN6 = 0</i>	<i>LIEN6 = 1</i>	Total
<i>Core - UE = 0</i> (en %)	3423 85	603 15	4026 100
<i>Core - UE = 1</i> (en %)	2158 89,4	257 10,6	2415 100

Tableau J : Réalisation des liens en fonction de l'appartenance des agents au cœur.

	<i>LIEN6 = 0</i>	<i>LIEN6 = 1</i>	Total
<i>Core1/2 = 0</i> (en %)	2940 87,5	421 12,5	3361 100
<i>Core1/2 = 1</i> (en %)	2641 85,7	439 14,3	3080 100

Tableau K : Réalisation des liens en fonction de l'appartenance d'un des deux agents au cœur.

Annexes A.15 : Statistiques descriptives dans le cas où $LIEN5 = 1$

Variables	Observations	Moyenne	Ecart-type	Minimum	Maximum
<i>Lien6</i>	306	0,41	0,49	0	1
<i>Lien5</i>	306	1,00	0,00	1	1
<i>Proxgeo</i>	306	0,42	0,50	0	1
<i>coreUE</i>	306	0,29	0,45	0	1
<i>COST</i>	306	514357,44	276222,69	77977,5	1384563,843
<i>$\Delta COST$</i>	306	417911,75	368321,26	1586	1551167,6
<i>Partcom</i>	306	4,88	4,14	0	18
<i>DR</i>	306	1111,89	150,36	1,5	1359,158
<i>ΔDR</i>	306	149,16	114,42	0	404,183

Annexes A.15 : Statistiques descriptives dans le cas où $LIEN5 = 0$

Variables	Observations	Moyenne	Ecart-type	Minimum	Maximum
<i>Lien6</i>	6135	0,12	0,32	0	1
<i>Lien5</i>	6135	0,00	0,00	0	0
<i>Proxgeo</i>	6135	0,37	0,48	0	1
<i>coreUE</i>	6135	0,38	0,49	0	1
<i>COST</i>	6135	439656,62	246703,69	33700	2028476,3
<i>$\Delta COST$</i>	6135	349699,38	355747,54	196	2465753
<i>Partcom</i>	6135	0,34	0,75	0	7
<i>DR</i>	6135	909,94	199,03	2,75	1307
<i>ΔDR</i>	6135	273,38	311,84	0	1372,35

Annexes A.15 : Les déterminants des collaborations : Prise en compte des liens antérieurs

	Modèle 13	Modèle 14	Modèle 15	Modèle 16 <i>LIEN5</i> = 1	Modèle 17 <i>LIEN5</i> = 0
<i>CONSTANTE</i>	-4,779*** (0,315)	-4,629*** (0,326)	-5,482*** (0,312)	-1,837* (0,329)	-4,923*** (0,335)
\overline{DR}	0,0023*** (0,0003)	0,0021*** (0,0003)	0,0031*** (0,0003)	ns	0,0025*** (0,0003)
ΔDR	0,0022*** (0,0002)	0,0022*** (0,0002)	0,0024*** (0,0002)	ns	0,0023*** (0,0002)
\overline{COUT}	0,534*** (0,206)	0,554*** (0,206)	0,520*** (0,201)	0,112* (0,625)	0,459*** (0,221)
$\Delta COUT$	-0,479*** (0,149)	-0,489*** (0,149)	-0,442*** (0,145)	ns	-0,471*** (0,159)
<i>PARTCOM</i>		ns			
<i>Core – UE</i>	-0,361*** (0,0856)	-0,362*** (0,0861)	-0,371*** (0,085)	ns	-0,373*** (0,0907)
<i>PROXGEO</i>	0,199** (0,0812)	0,202*** (0,081)	0,228*** (0,080)	0,783*** (0,265)	ns
<i>LIEN5</i>	1,467*** (0,134)	1,330*** (0,172)			
<i>PSEUDO – R</i> ²	0,072	0,072	0,050	0,055	0,045
R^2_{VZ}	0,121	0,122	0,086	0,121	0,075
Nombre d'observations	6441	6441	6441	306	6135
<i>LIEN6</i> = 1	860	860	860	125	735
<i>LIEN6</i> = 0	5581	5581	5581	181	5400

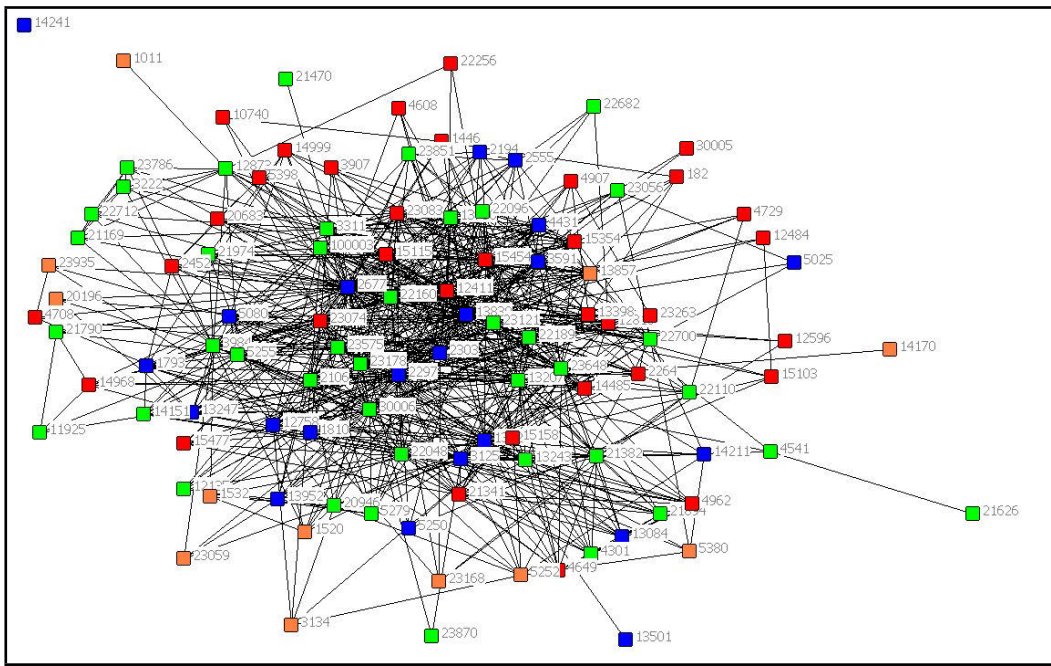
Tableau : Estimations des déterminants des choix de collaborations obtenues à partir d'un modèle Logit (6641 observations). Niveau de significativité : *0.10 ; ** 0.05 ; *** 0.01 ; ns = non significatif. Les écarts-types sont donnés entre parenthèses. Les valeurs des paramètres ainsi que celles des écarts-types associés sont arrondies à 10^{-2} et les *PSEUDO – R*² sont arrondis à 10^{-3} .

Annexes A.15 : Les déterminants des collaborations : Prise en compte des liens antérieurs

	<i>Lien6 = 0</i>	<i>Lien6 = 1</i>	Total
<i>Lien5 = 0</i> (en %)	5400 88,02	735 11,98	6135 100
<i>Lien5 = 1</i> (en %)	181 59,15	125 40,85	306 100

Tableau L : Réalisation des liens dans le 6^{ème} PCRDT en fonction des liens réalisés dans le 5^{ème} PCRDT.

Annexes A.16. Illustration des hypothèses



Graphe M : Réseaux des collaborations du 6^{ème} PCRDT. La couleur des points varient en fonction du type des agents. Les points rouges sont les entreprises non PME, les points oranges sont les entreprises PME, les points bleus sont les centres de recherche et les points verts sont les agents de l'enseignement supérieur.

Annexes A.16 : Hétérogénéité des agents et des comportements collaboratifs. Statistiques descriptives pour les collaborations entre centres de recherche et/ou agents de l'enseignement supérieur

Variable	Observations	Moyenne	Écart-type	Minimum	Maximum
<i>Lien6</i>	1953	0,20635	0,40479	0	1
<i>Lien5</i>	1953	0,06349	0,24391	0	1
\overline{COUT}	1953	344965	169968	44131	1142439
$\Delta COUT$	1953	251802	236451	196	1177928
\overline{DR}	1953	943,32283	196,8208	1,5	1307
ΔDR	1953	262,96412	301,41102	0	1343
<i>Partcom</i>	1953	0,76856	1,74845	0	13
<i>PROXGEO</i>	1953	0,29442	0,4559	0	1
<i>core – UE</i>	1953	0,32258	0,46758	0	1

Annexes A.16 : Hétérogénéité des agents et des comportements collaboratifs. Statistiques descriptives pour les collaborations mixtes

Variable	Observations	Moyenne	Écart-type	Minimum	Maximum
<i>Lien6</i>	2898	0,11111	0,31432	0	1
<i>Lien5</i>	2898	0,04279	0,20241	0	1
\overline{COUT}	2898	469098	244700	33700	1841857
$\Delta COUT$	2898	391715	384287	419,5	2465753
\overline{DR}	2898	915,81062	207,57993	2,75	1359
ΔDR	2898	277,40568	313,69724	0	1372
<i>Partcom</i>	2898	0,5559	1,53561	0	17
<i>PROXGEO</i>	2898	0,38923	0,48766	0	1
<i>core – UE</i>	2898	0,36025	0,48016	0	1

Annexes A.16 : Hétérogénéité des agents et des comportements collaboratifs. Statistiques descriptives pour les collaborations entre entreprises (PME ou non PME)

Variable	Observations	Moyenne	Écart-type	Minimum	Maximum
<i>Lien6</i>	1035	0,09758	0,2969	0	1
<i>Lien5</i>	1035	0,04444	0,20618	0	1
\overline{COUT}	1035	593231	297396	76589	2028476
$\Delta COUT$	1035	432360	427484	665	2438385
\overline{DR}	1035	888,2984	214,00966	4	1351
ΔDR	1035	291,81579	326,01999	0	1370
<i>Partcom</i>	1035	0,42319	1,24341	0	18
<i>PROXGEO</i>	1035	0,50242	0,50024	0	1
<i>core – UE</i>	1035	0,39227	0,48849	0	1

Annexes A.16 : Hétérogénéité des agents et des comportements collaboratifs

	Modèle 18	Modèle 19	Modèle 20	Modèle 21	Modèle 22	Modèle 23
<i>CONSTANTE</i>	-4,737*** (0,327)	-4,631*** (0,327)	-4,567*** (0,327)	-4,631*** (0,329)	-4,916*** (0,330)	-4,860*** (0,331)
\overline{DR}	0,0021*** (0,0003)	0,0021*** (0,0003)	0,0020*** (0,0003)	0,0019*** (0,0003)	0,0017*** (0,0003)	0,0017*** (0,0003)
ΔDR	0,0022*** (0,0002)	0,0021*** (0,0002)	0,0021*** (0,0002)	0,0021*** (0,0002)	0,0021*** (0,0002)	0,0021*** (0,0002)
\overline{COUT}	0,758*** (0,208)	0,595*** (0,204)	0,926*** (0,213)	0,100*** (0,218)	0,117*** (0,211)	0,129*** (0,179)
$\Delta COUT$	-0,489*** (0,146)	-0,486*** (0,147)	-0,587*** (0,148)	-0,540*** (0,148)	-0,466*** (0,147)	-0,508*** (0,148)
<i>PARTCOM</i>	0,157*** (0,0213)	0,163*** (0,0213)	0,161*** (0,0213)	0,159*** (0,0213)	0,154*** (0,0213)	0,154*** (0,0213)
<i>Core – UE</i>	-0,372*** (0,0856)	-0,366*** (0,0859)	-0,383*** (0,0859)	-0,379*** (0,0863)	-0,364*** (0,0824)	-0,369*** (0,0856)
<i>PROXGEO</i>	0,254*** (0,0812)	0,227*** (0,0810)	0,266*** (0,0814)	0,283*** (0,0821)	0,311*** (0,0824)	0,326*** (0,0827)
<i>HE</i>	0,372*** (0,11)			0,384*** (0,111)		
<i>RES</i>		0,827*** (0,159)		0,807***		
<i>IND</i>			-0,560*** (0,122)	-0,473*** (0,123)		-0,287** (0,127)
<i>HERES</i>					0,919*** (0,0832)	0,868*** (0,0856)
<i>PSEUDO – R²</i>	0,063	0,066	0,066	0,072	0,085	0,086
R^2_{VZ}	0,107	0,112	0,112	0,121	0,142	0,144
Nombre d'observations	6441	6441	6441	6441	6441	6441
<i>LIEN6 = 1</i>	860	860	860	860	860	860
<i>LIEN6 = 0</i>	5581	5581	5581	5581	5581	5581

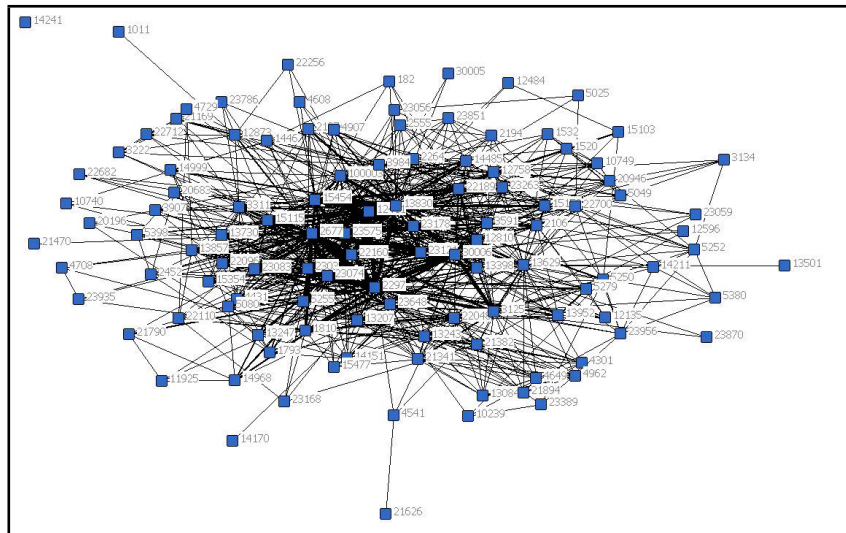
Tableau : Estimations des déterminants des choix de collaborations obtenues à partir d'un modèle Logit (5581 observations). Niveau de significativité : *0.10 ; ** 0.05 ; *** 0.01 ; ns = non significatif. Les écarts-types sont donnés entre parenthèses. Les valeurs des paramètres ainsi que celles des écarts-types associés sont arrondies à 10^{-2} et les *PSEUDO – R²* sont arrondis à 10^{-3} .

Annexes A.16 : Hétérogénéité des agents et des comportements collaboratifs

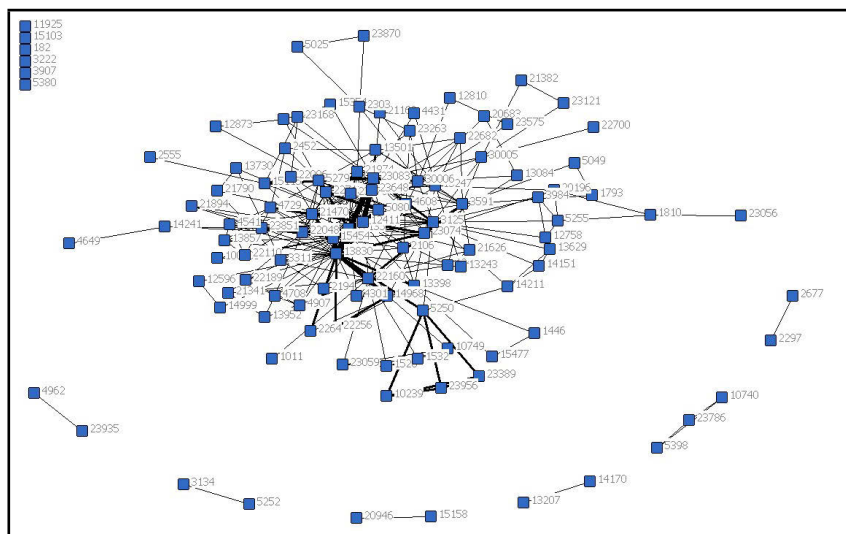
	Modèle 24	Modèle 25 <i>HERES</i>	Modèle 26 <i>MIX</i>	Modèle 27 <i>IND</i>
<i>CONSTANTE</i>	-4,750*** (0,337)	-3,187*** (0,53)	-4,835*** (0,514)	-4,631*** (0,329)
\overline{DR}	0,0023*** (0,0003)	0,0011** (0,0005)	0,0017*** (0,0005)	0,0044*** (0,0011)
ΔDR	0,0022*** (0,0002)	0,0020*** (0,0003)	0,0020*** (0,0003)	0,0021*** (0,0005)
\overline{COUT}	0,580*** (0,211)	ns	0,151*** (0,332)	ns
$\Delta COUT$	-0,497*** (0,152)	ns	-0,634*** (0,215)	ns
<i>PARTCOM</i>	0,145*** (0,0214)	0,076*** (0,337)	0,173*** (0,033)	0,550*** (0,112)
<i>Core – UE</i>	-0,366*** (0,0885)	-0,537*** (0,1398)	-0,281** (0,137)	ns
<i>PROXGEO</i>	0,166** (0,0825)	0,344*** (0,1319)	ns	ns
<i>PSEUDO – R²</i>	0,059	0,055	0,068	0,195
R^2_{VZ}	0,101	0,105	0,110	0,284
Nombre d'observations	5886	1953	2898	1035
<i>LIEN6 = 1</i>	826	403	322	101
<i>LIEN6 = 0</i>	5060	1550	2576	934

Tableau : Estimations des déterminants des choix de collaborations obtenues à partir d'un modèle Logit. Niveau de significativité : *0.10; ** 0.05; *** 0.01; ns = non significatif. Les écarts-types sont donnés entre parenthèses. Les valeurs des paramètres ainsi que celles des écarts-types associés sont arrondies à 10^{-2} et les *PSEUDO – R²* sont arrondis à 10^{-3} .

Annexes A.17. Illustration des hypothèses



Graphes N : Réseaux des collaborations du 6^{ème} PCRDT. L'épaisseur des traits varient en fonction de l'intensité des collaborations.



Graphes O : Réseaux des collaborations du 5^{ème} PCRDT. L'épaisseur des traits varient en fonction de l'intensité des collaborations.

Annexes A.17 : Intensité des liens

	Modèle 28	Modèle 29	Modèle 30	Modèle 31
<i>CONSTANTE</i>	-4,611*** (0,255)	-4,677*** (0,249)	-4,611*** (0,391)	-4,677*** (0,310)
\overline{DR}	0,0019*** (0,0003)	0,0019*** (0,0002)	0,0020*** (0,0004)	0,0019*** (0,0003)
ΔDR	0,0023*** (0,0001)	0,0023*** (0,0001)	0,0021*** (0,0002)	0,0021*** (0,0002)
\overline{COUT}	0,936*** (0,1431)	0,936*** (0,145)	0,936*** (0,219)	0,936*** (0,181)
$\Delta COUT$	-0,5654*** (0,1066)	-0,587*** (0,109)	-0,565*** (0,163)	-0,587*** (0,135)
<i>PARTCOM</i>	0,147*** (0,010)		0,147*** (0,016)	
<i>LIEN5</i>		1,352*** (0,089)		1,352*** (0,111)
<i>Core – UE</i>	-0,409*** (0,070)	-0,389*** (0,070)	-0,409*** (0,107)	-0,389*** (0,087)
<i>PROXGEO</i>	0,299*** (0,064)	0,275*** (0,064)	0,299*** (0,098)	0,275*** (0,080)
<i>Scale</i>	1 (0)	1 (0)	1,513 (0)	1,245 (0)
Nombre d'observations	6441	6441	6441	6441
<i>LIEN6 = 1</i>	860	860	860	860
<i>LIEN6 = 0</i>	5581	5581	5581	5581

Tableau : Estimations des déterminants des choix de collaborations obtenues à partir d'un modèle de Poisson (6441 observations). Niveau de significativité : *0.10 ; ** 0.05 ; *** 0.01 ; ns = non significatif. Les écarts-types sont donnés entre parenthèses. Les valeurs des paramètres ainsi que celles des écarts-types associés sont arrondies à 10^{-2} et les $PSEUDO - R^2$ sont arrondis à 10^{-3} .

Annexes A.17 : Intensité des liens

	Modèle 32	Modèle 33
<i>CONSTANTE</i>	-4,174*** (0,375)	-4,352*** (0,331)
\overline{DR}	0,0015*** (0,0004)	0,0017*** (0,0003)
ΔDR	0,0021*** (0,0002)	0,0021*** (0,0002)
\overline{COUT}	0,788*** (0,235)	0,8128*** (0,196)
$\Delta COUT$	-0,465*** (0,17)	-0,512*** (0,143)
<i>PARTCOM</i>	0,167*** (0,023)	
<i>LIEN5</i>		1,412*** (0,123)
<i>Core – UE</i>	-0,398*** (0,103)	-0,385*** (0,087)
<i>PROXGEO</i>	0,276*** (0,096)	0,236*** (0,081)
<i>Scale</i>	1,062 (0,199)	0,972 (0,156)
Nombre d'observations	6441	6441
<i>LIEN6 = 1</i>	860	860
<i>LIEN6 = 0</i>	5581	5581

Tableau : Estimations des déterminants des choix de collaborations obtenues à partir d'un modèle binomial négatif (6441 observations). Niveau de significativité : *0.10 ; ** 0.05 ; *** 0.01 ; ns = non significatif. Les écarts-types sont donnés entre parenthèses. Les valeurs des paramètres ainsi que celles des écarts-types associés sont arrondies à 10^{-2} et les $PSEUDO - R^2$ sont arrondis à 10^{-3} .

Annexes A.17 : Intensité des liens

<i>Lien5</i>	Effectif	Moyenne de l'intensité des liens dans le 6 ^{ème} PCRDT
0	6135	0,147
1	306	0,660

Tableau M : Intensité des liens dans le 6^{ème} PCRDT en fonction de l'existence d'un lien dans le 5^{ème} PCRDT.

Intensité des liens	Observations	Moyenne \overline{Cost} (Ecart-type)	Moyenne $\Delta Cost$ (Ecart-type)
0	5581	440099 (247324)	352477 (362145)
1	717	439962 (245068)	339385 (311666)
2	100	568897 (296731)	432248 (338463)
3	20	550506 (244115)	428380 (368955)
4	11	685204 (120481)	681886 (268532)
5	3	474275 (231227)	238858 (227461)
6	5	589169 (281582)	214399 (291213)
8	2	909487 (253347)	576915 (506695)
9	1	375000 (-)	224516 (-)
10	1	1040241 (-)	315406 (-)

Tableau N : Valeur des variables liées aux moyens en R&D en fonction de l'intensité des collaborations.

Annexes A.17 : Intensité des liens

<i>Proxgeo</i>	Observations	Intensité moyenne
0	4028	0,148
1	2413	0,209

Tableau M : Intensité des liens dans le 6^{ème} PCRDТ en fonction de la proximité géographique mesurée par la variable *Proxgeo*.

Annexes A.18 : Formation de triades

	Moyenne	Ecart-type	Minimum	Maximum	Somme
$Triade6$	0,03	0,17	0	1	7111
\overline{COUT}^T	0,44	0,20	0,0	1,8	106575,0
$\Delta COUT^T$	0,53	0,40	0,0	2,5	127304,1
\overline{DR}^T	919,54	163,87	2,3	1348,7	221115779,2
ΔDR^T	401,21	351,63	0	1372,4	96476883,7
$PARTCOM^T$	0,05	0,16	0	4,3	13029,3
$PROXGEO3^T$	0,13	0,34	0	1	32295
$PROXGEO2^T$	0,14	0,35	0	1	34357
$Core - UE3^T$	0,23	0,42	0	1	54740
$Core - UE2^T$	0	0	0	0	0
$Core - UE1^T$	0,44	0,50	0	1	106260
\overline{DIST}^T	1178,80	569,02	0	3845	283459642,7
$\Delta DIST^T$	1005,51	617,55	0	4253	241788004
\overline{TEMPS}^T	755,51	393,38	0	2910,7	181672512
$\Delta TEMPS^T$	643,62	421,32	0	3071	154767570
$Pays^T$	0,13	0,33	0	1	30906
TC^T	0,01	0,11	0	1	3090

Tableau : Statistiques descriptives des variables portant sur les triades.

Annexes A.18 : Formation de triades

	Modèle 34	Modèle 35	Modèle 36	Modèle 37	Modèle 38	Modèle 39
$CONSTANTE$	-4,81*** (0,17)	-4,81*** (0,17)	-4,81*** (0,17)	-4,81*** (0,17)	-4,63*** (0,18)	-4,46*** (0,17)
\overline{DR}^T	0,0006*** (0,0002)	0,0006*** (0,0002)	0,0006*** (0,0002)	0,0006*** (0,0002)	0,0006*** (0,0002)	0,0005*** (0,0002)
ΔDR^T	0,0012*** (0,0001)	0,0012*** (0,0001)	0,0012*** (0,0001)	0,0012*** (0,0001)	0,0012*** (0,0001)	0,0012*** (0,0001)
\overline{COUT}^T	ns	ns	ns	ns	ns	ns
$\Delta COUT^T$	ns	ns	ns	ns	ns	ns
$PARTCOM^T$	0,98*** (0,07)	0,98*** (0,07)	0,98*** (0,07)	0,97*** (0,08)	0,97*** (0,07)	0,98*** (0,08)
$PROXGEO3^T$		0,12*** (0,05)		0,13*** (0,05)		
$PROXGEO2^T$			ns	ns		
\overline{TEMPS}^T					-0,0003*** (0,0001)	
$\Delta TEMPS^T$					ns	
\overline{DIST}^T						-0,0002*** (0,0001)
$\Delta DIST^T$						ns
$PSEUDO - R^2$	0,029	0,029	0,029	0,029	0,029	0,03
R_{VZ}^2	0,036	0,037	0,036	0,037	0,037	0,038

Tableau : Estimations des déterminants des choix de collaborations obtenues à partir d'un modèle logit négatif (6441 observations). Niveau de significativité : *0.10; ** 0.05; *** 0.01; ns = non significatif. Les écarts-types sont donnés entre parenthèses. Les valeurs des paramètres ainsi que celles des écarts-types associés sont arrondies à 10^{-2} et les $PSEUDO - R^2$ sont arrondis à 10^{-3} .

Annexes A.18 : Formation de triades

	Modèle 40	Modèle 41	Modèle 42	Modèle 43	Modèle 44	Modèle 45	Modèle 46
$CONSTANTE$	-4,70*** (0,17)	-4,77*** (0,17)	-4,85*** (0,17)	-4,47*** (0,18)	-4,81*** (0,17)	-5,54*** (0,16)	-4,96*** (0,17)
\overline{DR}^T	0,005*** (0,0002)	0,006*** (0,0002)	0,005*** (0,0002)	0,005*** (0,0002)	0,005*** (0,0002)	0,0012*** (0,0005)	0,006*** (0,0002)
ΔDR^T	0,0012*** (0,0001)	0,0012*** (0,0001)	0,0012*** (0,0001)	0,0012*** (0,0001)	0,0014*** (0,0001)	(0,0002)	0,0013*** (0,0001)
\overline{COUT}^T	ns	ns	ns	ns	ns	ns	ns
$\Delta COUT^T$	ns	ns	ns	ns	ns	ns	ns
$PARTCOM^T$	0,96*** (0,008)	0,96*** (0,008)	0,96*** (0,008)	0,95*** (0,008)	0,88*** (0,008)		0,99*** (0,008)
$PROXGEO3^T$	0,18*** (0,05)	0,13*** (0,05)	0,16*** (0,05)	0,23*** (0,05)	0,12*** (0,05)	0,16*** (0,04)	
$CORE - UE3^T$	-0,36*** (0,05)			-0,56*** (0,07)			
$CORE - UE2^T$		-0,07** (0,04)		-0,29*** (0,06)			
$CORE - UE1^T$			0,20*** (0,04)	ns			
TC^T					0,33*** (0,1)	0,79*** (0,09)	
$Pays^T$							0,28*** (0,05)
$PSEUDO - R^2$	0,031	0,030	0,032	0,032	0,030		0,030
R_{VZ}^2	0,039	0,038	0,040	0,040	0,037		0,038

Tableau : Estimations des déterminants des choix de collaborations obtenues à partir d'un modèle logit négatif (6441 observations). Niveau de significativité : *0.10 ; ** 0.05 ; *** 0.01 ; ns = non significatif. Les écarts-types sont donnés entre parenthèses. Les valeurs des paramètres ainsi que celles des écarts-types associés sont arrondies à 10^{-2} et les $PSEUDO - R^2$ sont arrondis à 10^{-3} .